

Simulating Branching Programs with Edit Distance and Friends Or: A Polylog Shaved is a Lower Bound Made*

Amir Abboud
Stanford University
abboud@cs.stanford.edu

Thomas Dueholm Hansen
Aarhus University
tdh@cs.au.dk

Virginia Vassilevska Williams
Stanford University
virgi@cs.stanford.edu

Ryan Williams
Stanford University
rrwilliams@gmail.com

Abstract

A recent and active line of work achieves tight lower bounds for fundamental problems under the Strong Exponential Time Hypothesis (SETH). A celebrated result of Backurs and Indyk (STOC'15) proves that the Edit Distance of two sequences of length n cannot be computed in strongly subquadratic $O(n^{2-\varepsilon})$ time, for some $\varepsilon > 0$, under SETH. Follow-up works proved that even simpler looking problems like the Longest Common Subsequence or Edit Distance of *binary* sequences suffer from the same SETH lower bound.

SETH is a very strong assumption, asserting that even *linear* size CNF formulas cannot be analyzed for satisfiability with an exponential speedup over exhaustive search. We consider much safer assumptions, e.g. that SAT on more expressive representations, like subexponential-size NC circuits, cannot be solved in $(2 - \delta)^n$ time. Intuitively, this assumption is much more plausible: NC circuits can implement linear algebra and complex cryptographic primitives, while CNFs cannot even approximately compute an XOR of bits.

Our main result is a surprising reduction from SAT on Branching Programs to fundamental problems in P like Edit Distance, LCS, and many others. A consequence of our reduction is that truly subquadratic algorithms will have consequences that we consider to be far more remarkable than merely faster CNF SAT algorithms. For example: SAT on arbitrary nondeterministic branching programs of size $2^{o(\sqrt{n})}$ or on arbitrary $o(n)$ -depth bounded fan-in circuits (and therefore also NC-Circuit-SAT) can be solved in $(2 - \delta)^n$ time. To illustrate the significance of these consequences we point at several new circuit lower bounds, that are very far from the state of the art, which immediately follow.

A very interesting feature of our work is that we can prove major consequences even from *mildly* subquadratic algorithms for Edit Distance or LCS. For example, we show that if we can shave an arbitrarily large polylog factor from n^2 for Edit Distance then NEXP does not have non-uniform NC¹ circuits. A more fine-grained examination shows that even shaving a $\log^c n$ factor, for a specific constant $c \approx 10^3$, already implies new circuit lower bounds.

*Part of the work was performed while visiting the Simons Institute for the Theory of Computing, Berkeley, CA. A.A. and V.V.W. were supported by NSF Grants CCF-1417238 and CCF-1514339, and BSF Grant BSF:2012338. T.D.H. was supported by the Carlsberg Foundation, grant no. CF14-0617. R.W. was supported by an Alfred P. Sloan Fellowship and NSF CCF-1212372. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

1 Introduction

A central goal of complexity theory is to understand and prove lower bounds for the time complexity of fundamental problems. One of the most important computational problems is Edit-Distance, the problem of computing the minimum number of edit operations (insertions, deletions, and substitutions) required to transform one sequence into another. A classical dynamic programming algorithm that is taught in basic algorithms courses solves Edit-Distance on sequences of length n in $O(n^2)$ time [29]. This quadratic runtime is prohibitive in many applications, like computational biology and genomics where n is typically a few billions. The significance of a faster, e.g. linear time, algorithm cannot be overstated. Despite decades of attempts, no upper bound below $O(n^2/\log^2 n)$ is known for Edit-Distance [35]. All the above applies to the simpler looking Longest Common Subsequence problem (LCS), for which the existence of a faster algorithm was already posed as an important open question in combinatorics by Knuth decades ago [28]. This is a situation where lower bounds are highly desirable, but unfortunately, the current state of the art in complexity is far from providing a lower bound that is close to quadratic for any natural problem in NP, let alone Edit-Distance. Therefore, researchers have turned their attention to conditional lower bounds, and a recent breakthrough by Backurs and Indyk [16] showed that Edit-Distance cannot be solved in truly subquadratic $O(n^{2-\varepsilon})$ time, for some $\varepsilon > 0$, unless the following well-known hypothesis on the complexity of k -SAT is false.

Hypothesis 1 (Strong exponential time hypothesis (SETH)). *There does not exist an $\varepsilon > 0$ such that for all $k \geq 3$, k -SAT on n variables and m clauses can be solved in $O(2^{(1-\varepsilon)n} \cdot m)$ time.*

This lower bound was received with a great deal of excitement¹. Other interesting recent results show that under SETH, the current algorithms for many central problems in diverse areas of computer science are optimal, up to $n^{o(1)}$ factors. These areas include pattern matching [10, 7, 22, 6], graph algorithms [39, 8, 9, 11, 13], parameterized complexity [36, 7], computational geometry [21, 23], and the list is growing by the day. Bringmann and Künnemann [22] generalize many of the previous SETH lower bounds [10, 21, 16, 7] into one framework; they prove that the problem of computing any similarity measure δ over two sequences (of bits, symbols, points, etc) will require quadratic time, as long as the similarity measure has a certain property (namely, if δ admits *alignment gadgets*). Such similarity measures include Edit-Distance and LCS (even on *binary* sequences), and the Dynamic Time Warping Distance, which is an extensively studied metric in time-series analysis.

These SETH lower bounds are a part of a more general line of work in which one bases the hardness of important problems in P on well-known conjectures about the exact complexity of other famous problems. Other conjectures are 3-SUM and All-Pairs-Shortest-Paths, but in recent years, SETH has been the most “successful” conjecture at explaining barriers.

How strong is SETH? SETH was introduced [31, 25] as a plausible explanation for the lack of $(2 - \varepsilon)^n$ algorithms for CNF-SAT, despite of it being one of the most extensively studied problems in computer science. The fastest known algorithms for k -SAT run in time $2^{n-n/O(k)}$ (e.g. [37]), and for CNF-SAT the bound is $2^{n-n/O(\log \Delta)}$ where $\Delta = m/n$ is the clause-to-variable ratio [24, 30, 12]. That is, these algorithms are *just* not enough to refute SETH. Evidence in favor of SETH is

¹It is among the few STOC papers that made it to the Boston Globe news website [1]. See also the Quanta article [2], the MIT news article [3], and the blog posts by Aaronson [4] and Lipton [5].

circumstantial. For example, natural algorithmic approaches like resolution were shown to require exponentially many steps [19].

There is evidence that SETH will be hard to *refute*, in the form of a “*circuit lower bounds barrier*”: refuting SETH is as hard as proving longstanding open lower bound results. Williams showed that faster-than-trivial Circuit-SAT algorithms for many circuit classes \mathcal{C} would imply interesting new lower bounds against that class [44, 46]. Via this connection, and known reductions from certain circuit families to CNF formulas, it is possible to show that refuting SETH implies a new circuit lower bound [33]: E^{NP} cannot be solved by *linear-size series-parallel circuits*². However, this is a *very* weak lower bound consequence.

A hierarchy of SAT problems. A weakness of SETH as a hardness hypothesis is that it is an assumption about CNF SAT, as opposed to a more general SAT problem. Consider a problem in which you are given some representation of a function \mathcal{B} on n input bits and are asked whether \mathcal{B} is satisfiable. If \mathcal{B} is treated as a black-box that we only have input/output access to, then any algorithm will need to spend $\Omega(2^n)$ time in the worst case. Of course, a clever algorithm should attempt to analyze \mathcal{B} in order to decide satisfiability in $o(2^n)$ time. Whether this is possible, depends on how complex and obfuscating the representation is. There is a whole spectrum of increasing complexity of representations, starting from simple objects like DNFs, which are very bad at hiding their satisfiability, up until large circuits or nondeterministic turing machines that we have no idea how to analyze.

For each class of representations \mathcal{C} we can define the corresponding \mathcal{C} -SETH, stating that this abstract SAT problem cannot be solved in $(2 - \varepsilon)^n$ time. For example, NC-SETH would be the assumption that Circuit-SAT on polynomial size polylog depth circuits (NC circuits) cannot be solved in $(2 - \varepsilon)^n$ time. It is well known that NC circuits are capable of complex computations, including most linear-algebraic operations. Moreover, they are believed to be capable of implementing cryptographic primitives like One Way Functions and Pseudorandom Generators, for which the ability to hide satisfiability is essential. In sharp contrast, the original SETH is equivalent (due to the sparsification lemma [32, 24]) to the assumption that even representations that are very low on this spectrum, namely *linear size CNF formulas*, are enough to obfuscate satisfiability. While from the viewpoint of polynomial time solvability, CNF-SAT and NC-SAT are equivalent, this is not the case from a more fine-grained perspective: an algorithm that can decide satisfiability of arbitrary *polynomial* size circuits faster than exhaustive search is far more remarkable than a similar algorithm that can handle only linear size CNF formulas.

As our class \mathcal{C} gets more complex and rich, the \mathcal{C} -SETH becomes more credible and appealing as a basis for conditional lower bounds than SETH. However, all previous SETH lower bound proofs relied heavily on the simple nature of CNFs. In this work, we prove the first lower bounds under the much more reliable \mathcal{C} -SETH, for classes \mathcal{C} that are far more expressive than CNFs.

Our results. Our main result is a new efficient reduction from SAT on super-polynomial size *nondeterministic Branching Programs* (BPs) to Edit-Distance, LCS and many other important problems in P. As we discuss below, BPs are vastly more expressive than CNFs. For example, our reduction allows us to *tightly* reduce SAT on arbitrary polynomial size NC circuits to problems

²The class E^{NP} or $\text{TIME}[2^{O(n)}]^{\text{NP}}$ is the class of problems solvable in exponential time with access to an NP oracle. Series-parallel circuits are a special kind of log-depth circuits, also known as Valiant-Series-Parallel circuits [42].

in P. Thus, we are able to replace SETH with NC-SETH, and derive far more remarkable consequences from truly subquadratic algorithms for Edit Distance and LCS. Moreover, we show that *any* problem for which the general framework of Bringmann and Künnemann is capable of showing an $n^{2-o(1)}$ SETH lower bound, will suffer from the much stronger NC-SETH hardness barrier. In fact, we are able to show reductions even to problems that fall outside their framework, like LCS on k sequences, a classical problem in parameterized complexity with an $O(n^k)$ algorithm [20, 38, 7].

BPs are a popular non-uniform model of computation [15]. Roughly speaking, a nondeterministic Branching Program on n input variables of width W and length T is a layered directed graph on T layers, each layer having W nodes, such that every edge is labelled with a constraint of the form $(x_i = b)$ where x_i is an input variable, and $b \in \{0, 1\}$. Note that typically $T \gg n$ and each node appears on many edges. One of the nodes in the first layer is marked as the *start node*, and one of the nodes in the last layer is marked as the *accept node*. For an input $x \in \{0, 1\}^n$ let G_x be the subgraph of edges whose constraints are satisfied by x . We say that the BP accepts the input x iff the accept node is reachable from the start node in G_x . The *size* of a BP is the total number of edges, i.e. $O(W^2T)$. We refer to Section 2 for additional details about branching programs. Even when the width is constant, BPs are surprisingly expressive: Barrington’s Theorem states that any fan-in 2, depth d circuit can be converted into an equivalent BP of width 5 and size 4^d , over the same set of inputs [18]. Therefore, any circuit with fan-in 2 of polylog depth of any size (in particular, NC circuits) can be expressed as a BP of length $2^{\text{polylog } n}$ and constant width. Our reduction shows that truly subquadratic Edit Distance would imply a $(2 - \delta)^n$ algorithm for SAT of constant-width $2^{o(n)}$ -length BPs.

Theorem 2. *There is a reduction from SAT on nondeterministic branching programs on n variables, length T , and width W , to an instance of Edit-Distance or LCS on two binary sequences of length $N = 2^{n/2} \cdot T^{O(\log W)}$, and the reduction runs in $O(N)$ time.*

Besides the constant width case, another interesting setting is where W and T are both $2^{o(\sqrt{n})}$, which corresponds to BPs that can represent *any nondeterministic Turing machine* that uses $o(\sqrt{n})$ space [15]. Thus, truly subquadratic Edit Distance or LCS would allow us to get an exponential improvement over exhaustive search for checking SAT of complex objects that can easily implement cryptographic primitives, and many of our favorite algorithms. This would be much more surprising than a faster SAT algorithm on linear size CNFs (as in SETH). To support this, we point at a few strong circuit lower bounds that would follow from such an algorithm.

If we assume that Edit-Distance or LCS can be solved in truly subquadratic time, then (among other things) Theorem 2 implies $O(2^{n-\varepsilon n/2})$ time algorithms for SAT on arbitrary formulas of size $2^{o(n)}$ and for SAT on nondeterministic branching programs of size $2^{o(\sqrt{n})}$. Combining this with connections between faster SAT algorithms and circuit lower bounds from prior work (see Section 6 for formal statements and a sketch of the proof), we obtain the following circuit lower bound consequences.

Corollary 1. *If Edit Distance or LCS on two binary sequences of length N is in $O(N^{2-\varepsilon})$ time for some $\varepsilon > 0$, then the complexity class \mathbf{E}^{NP} does not have:*

1. *non-uniform $2^{o(n)}$ -size Boolean formulas,*
2. *non-uniform $o(n)$ -depth circuits of bounded fan-in, and*
3. *non-uniform $2^{o(n^{1/2})}$ -size nondeterministic branching programs.*

Furthermore, $\text{NTIME}[2^{O(n)}]$ is not in non-uniform NC.

The above lower bound consequences are far stronger than any state of the art. The first consequence is interesting due to the rarity of $2^{\Omega(n)}$ circuit lower bounds: it is still open whether the humongous class $\Sigma_2\text{EXP}$ has $2^{o(n)}$ size *depth-three* circuits. The third consequence is interesting because it yields an exponential lower bound for arbitrary nondeterministic BPs; this model is vastly bigger than NL/poly . The fourth is interesting because the lower bound holds for the smaller class $\text{NTIME}[2^{O(n)}]$. These consequences are on a different scale compared to the ones obtained from refuting SETH, and therefore the “circuit lower bounds barrier” for faster Edit Distance is much stronger.

Our first corollary was a strict improvement over the previously known SETH lower bounds, in terms of the significance of the consequences. Next, we show that our reduction allows us to derive consequences even from *mildly* subquadratic algorithms, a feature that did not exist in the previous conditional lower bounds in P.

Given the status of Edit-Distance and LCS as core computer science problems, any asymptotic improvement over the longstanding $O(n^2/\log^2 n)$ upper bound is highly desirable. Recent algorithmic techniques were able to beat similar longstanding bounds for other core problems like All Pairs Shortest Path (APSP) [45, 27], 3-SUM [34], and Boolean Matrix Multiplication [17, 26, 47]. For example, the *polynomial method* [45] has allowed for superpolylogarithmic shavings for APSP, and more recently to two other problems that are more closely related to ours, namely Longest Common Substring [12], and Hamming Nearest Neighbors [14]. A natural open question [45, 12, 14] is whether these techniques can lead to an $n^2/\log^{\omega(1)} n$ algorithm for Edit-Distance as well. The lower bound of Backurs and Indyk is not sufficient to address this question, and only a much faster $n^2/2^{\omega(\log n/\log \log n)}$ would have been required to improve the current CNF-SAT algorithms. Our approach of considering \mathcal{C} -SETH for more expressive classes \mathcal{C} allows us to prove strong “circuit lower bounds barriers” *even for shaving log factors*.

Any formula of size $O(n^f)$ can be transformed into an equivalent BP of width 5 and size $O(n^{8f})$ (first rebalance into a formula of depth $4f \log n$ [41] and then use Barrington’s Theorem [18]). Applying Theorem 1 to the resulting BPs, we get LCS instances of size $N = O(2^{n/2} \cdot n^{8fd})$, for some constant $d \leq 25$ (the constant depends on the problem and the alphabet size). Shaving an $\Omega((\log N)^{8fd+f+10})$ factor would translate into an $O(2^n/(n^{10} \cdot n^f))$ algorithm for SAT of formulas of size $O(n^f)$. Thus, we get that if LCS can be solved in $O(n^2/\log^{1000} n)$ time, then SAT on formulas of size $O(n^5)$ can be solved in $O(2^n/n^{15})$ time, which would imply that E^{NP} does not have such formulas. We obtain that solving Edit-Distance or LCS in $n^2/\log^{\omega(1)} n$ time still implies a major circuit lower bound, namely that $\text{NTIME}[2^{O(n)}]$ is not in non-uniform NC^1 .

Corollary 2. *If Edit Distance or LCS on two binary sequences of length N can be solved in $O(n^2/\log^c n)$ time for every $c > 0$, then $\text{NTIME}[2^{O(n)}]$ does not have non-uniform polynomial-size log-depth circuits.*

It is likely that these connections could be sharpened even further and that similar consequences can be derived even from shaving fewer log factors. Some inefficiencies in these connections are due to constructions of certain gadgets in our proofs, while others come from the framework for obtaining circuit lower bounds from faster SAT algorithms, and the reductions from circuits to BPs.

One striking interpretation of these corollaries is that when an undergraduate student learns the simple dynamic programming algorithms for Edit-Distance or Longest Common Subsequence and

wonders whether there is a faster algorithm, he or she is implicitly trying to resolve very difficult open questions in complexity theory.

Technical remarks. All known SETH lower bound proofs for problems in P have relied as a first step on a reduction [43] to the following *Orthogonal Vectors* (OV) problem: given a set of n boolean vectors $S \subseteq \{0,1\}^d$ of dimension $d = \omega(\log n)$, does there exist a pair $a, b \in S$ such that for all $j \in [d]$ we have $(a[j] = 0 \text{ or } b[j] = 0)$, i.e. the vectors are orthogonal or “disjoint”. If OV can be solved in $O(n^{2-\varepsilon})$ time, then CNF-SAT can be solved in $O(2^{(1-\varepsilon/2)n})$ time. It is important to notice that reductions in the direction are not known, i.e. refuting SETH is not known to imply subquadratic algorithms for some hard quadratic-time problems. Therefore, lower bounds under the assumption that OV requires $n^{2-o(1)}$ time are more reliable than SETH lower bounds. However, the above weaknesses of SETH apply to OV as well: a much harder problem is the \mathcal{C} -Satisfying-Pair problem, where instead of searching for an orthogonal pair of vectors, we ask for a pair of vectors that (together) satisfy a certain function that can be represented in more complex ways than an orthogonality check. Again, there is a spectrum of increasing expressiveness, and OV is quite low on it. Indeed, we have no idea how to solve the NC-Satisfying-Pair problem in $O(n^2/\log^3 n)$ time (it would readily imply faster NC-SAT algorithms), while for OV the current upper bound $n^{2-1/O(\log(d/\log n))}$ is barely not truly subquadratic. All the reductions in this paper (except for the k -LCS proof) are via a certain Branching-Program-Satisfying-Pair problem, which can be solved in quadratic time, while faster algorithms would be very surprising and imply all the aforementioned consequences.

Previous SETH lower bound proofs, when stripped of all the gadgetry, are rather simple, due to the simplicity of the OV problem (which, in turn, is due to the simplicity of CNFs). Each vector is represented by some *vector gadget*, so that two gadgets “align well” if and only if the vectors are *good* (in this case, orthogonal), and then all the gadgets are combined so that the “total score” reflects the existence of a good pair. Vector gadgets that are capable of checking orthogonality can be constructed in natural ways by concatenating *coordinate gadgets* that have straightforward functionality (checking that not both coordinates are 1), which in turn can be constructed via certain atomic sequences of constant size. We observe that these reductions do not exhaust the expressive capabilities of problems like Edit Distance and LCS.

Our new reductions follow this same scheme, except that the functionality of the vector gadgets is no longer so simple. Our main technical contribution is the construction of certain *reachability gadgets*, from which our vector gadgets are constructed. These gadgets are capable of checking reachability between two nodes in a subgraph (e.g. u_{start} and u_{acc}) of a graph (the branching program) that is obtained from two given vectors. These gadgets exhibit the ability of sequence similarity measures to execute nontrivial algorithmic tasks. Our reduction can be viewed as encoding of graphs into two sequences such that the optimal LCS must implicitly execute the classical small-space algorithm for directed reachability of Savitch’s Theorem [40].

Previous work on basing hardness on better hypotheses. Finding more reliable hypotheses (than SETH, 3-SUM, APSP, etc) that can serve as an alternative basis for the “hardness in P” is an important goal. Previous progress towards this end was achieved by Abboud, Vassilevska Williams, and Yu [11] where the authors prove tight lower bounds for various graph problems under the hypothesis that *at least one* of the SETH, APSP, and 3-SUM conjectures is true. The \mathcal{C} -SETH hypothesis (say, for $\mathcal{C} = \text{NC}$) that we consider in this work is incomparable in strength to

theirs, yet it has certain advantages. First, the known connections between faster SAT algorithms and circuit lower bounds allow us to point at remarkable consequences of refuting our hypothesis, which is not known for any of the previous conjectures. Second, it allows us to show barriers even for *mildly* subquadratic algorithms. And third, it allows us to explain the barriers for many problems like Edit Distance and LCS for which a lower bound under 3-SUM or APSP is not known (unless the alphabet size is near-linear [10]).

Organization of the paper. The rest of the paper is organized as follows. In Section 2 we define the SAT problem on Branching Programs (BP-SAT), and briefly describe how it is used as the source of our reductions. In Section 3 we give a direct and simplified reduction from BP-SAT to LCS and k -LCS. We present the framework of Bringmann and Künnemann [22] in Section 4, along with a sketch of our full reduction. We then present the details of the full reduction in Section 5. The full reduction also applies to LCS, and is more efficient than the simplified reduction from Section 3. In Section 6 we discuss the consequences of a faster algorithm for BP-SAT which follow from combining classical connections between formulas, low-depth circuits, and BPs, with the more modern connections between faster SAT algorithms and circuit lower bounds.

2 Satisfiability of Branching Programs

In this section we define the SAT problem on Branching Programs (BP-SAT), which we later reduce to various sequence-problems such as Edit Distance and LCS.

A nondeterministic *Branching Program* (BP) of length T and width W on n boolean inputs x_1, \dots, x_n is a layered directed graph P with T layers L_1, \dots, L_T . The nodes of P have the form (i, j) where $i \in [T]$ is the layer number and $j \in [W]$ is the index of the node inside the layer. The node $u_{\text{start}} = (1, 1)$ is called the starting node of the program, and the node $u_{\text{acc}} = (T, 1)$ is the accepting node of the program. For all layers $i < T$ except the last one: all nodes in L_i are marked with the same variable $x(i) = x_{f(i)}$, and each node has an arbitrary number of outgoing edges, each edge marked with 0 or 1. Note that typically $T \gg n$ and each variable appears in many layers.

An *evaluation* of a branching program P on an input $x_1, \dots, x_n \in \{0, 1\}$ is a path that starts at u_{start} and then (nondeterministically) follows an edge out of the current node: if the node is in layer L_i we check the value of the corresponding variable $x_{f(i)}$, denote it by $\eta \in \{0, 1\}$, and then we follow one of the outgoing edges marked with η . The BP *accepts* the input iff the evaluation path ends in u_{acc} . That is, each input restricts the set of edges that can be taken, and the BP accepts an input iff there is a path from u_{start} to u_{acc} in the subgraph induced by the input.

Definition 1 (BP-SAT). *Given a Branching Program P on n boolean inputs, decide if there is an assignment to the variables that makes P accept.*

To prove a reduction from BP-SAT to a sequence-problem we go through the following problem: Let $X_1 = \{x_1, \dots, x_{n/2}\}$ and $X_2 = \{x_{n/2+1}, \dots, x_n\}$ be the first and last half of the inputs to the branching program, respectively. Do there exist $a \in \{0, 1\}^{n/2}$ and $b \in \{0, 1\}^{n/2}$, such that when viewed as partial assignments to X_1 and X_2 , respectively, together they form an accepting input to the branching program? This problem is clearly just a reformulation of BP-SAT. Our reductions also work, however, when a and b are restricted to two given sets of vectors $S_1, S_2 \subseteq \{0, 1\}^{n/2}$, i.e., we ask whether there exists an accepting pair $(a, b) \in S_1 \times S_2$. We refer to this problem as the *satisfying pair problem* on branching programs. Proving a reduction from this more general problem

corresponds to proving a reduction from the *orthogonal vectors problem* (OV) to get a SETH-based lower bound (see [43]). To simplify the presentation we assume, however, that $S_1 = S_2 = \{0, 1\}^{n/2}$.

Our reductions construct for each set S_i a sequence composed of subsequences that correspond to elements of S_i . The length of these subsequences depends on the branching program but will generally be bounded by $2^{o(n)}$. The combined sequence-length is therefore $N = 2^{(1/2+o(1))n}$. This establishes a connection between BP-SAT, which is solvable in exponential time, and sequence-problems that are solvable in quadratic time.

3 A Simplified Reduction to Longest Common Subsequence

Given two strings of N symbols over some alphabet Σ , the longest common subsequence (LCS) problem asks for the length of the longest sequence that appears as a subsequence in both input strings. In this section we prove the following reduction from LCS to BP-SAT. To simplify the presentation we give a less efficient reduction that uses $|\Sigma| = O(W \log T)$ symbols for branching programs of width W and length T . We refer to sections 4 and 5 for a more efficient reduction with $|\Sigma| = 2$, that is based on the framework of Bringmann and Künnemann [22].

Theorem 3. *There is a constant c such that if LCS can be solved in time $S(N)$, then BP-SAT on n variables and programs of length T and width W can be solved in $S(2^{n/2} \cdot T^{c \log W})$ time.*

Let P be a given branching program on n boolean inputs, and let F be the corresponding function. As mentioned in Section 2, we prove Theorem 3 by reducing the satisfying pair problem on branching programs to LCS. Let therefore $X_1 = \{x_1, \dots, x_{n/2}\}$ and $X_2 = \{x_{n/2+1}, \dots, x_n\}$ be the first and last half of the inputs to F , respectively. For two partial assignments a and b in $\{0, 1\}^{n/2}$, we use the notation $a \odot b$ to denote concatenation, forming a complete assignment. We must decide whether there exist $a, b \in \{0, 1\}^{n/2}$ such that $F(a \odot b) = 1$.

For two sequences x, y , let $LCS(x, y)$ denote the length of the longest common subsequence of x and y . The reduction consists of two steps. First, for each $a \in \{0, 1\}^{n/2}$ we construct a sequence $G(a)$, and for each $b \in \{0, 1\}^{n/2}$ we construct another sequence $\overline{G}(b)$. These sequences are defined by recursive gadget constructions. Let Y be some integer that depends on the width W and length T of the branching program. The sequences are constructed such that $LCS(G(a), \overline{G}(b)) = Y$ if $F(a \odot b) = 1$, and such that $LCS(G(a), \overline{G}(b)) \leq Y - 1$ otherwise. Solving LCS for $G(a)$ and $\overline{G}(b)$ can therefore be viewed as evaluating $F(a \odot b)$. Constructing $G(a)$ and $\overline{G}(b)$ is the main challenge in proving our reduction. In previous such reductions from OV this step was nearly trivial.

The second step is to combine $G(a)$ for all $a \in \{0, 1\}^{n/2}$ into a single sequence A , and $\overline{G}(b)$ for all $b \in \{0, 1\}^{n/2}$ into a single sequence B . Let E be some integer that depends on the width W and length T of the branching program. Then A and B are constructed such that $LCS(A, B) = E$ if there exist $a, b \in \{0, 1\}^{n/2}$ with $F(a \odot b) = 1$, and such that $LCS(A, B) \leq E - 1$ otherwise. The construction of A and B therefore completes the reduction.

For the second step we simply use the following lemma by Abboud *et al.* [7]. The proof of the lemma uses *normalized vector gadgets* similar to those used in the reduction by Backurs and Indyk [16] from orthogonal vectors to Edit Distance. We sketch the proof in Section 4 in the context of Bringmann and Künnemann's framework [22], and give a formal proof of a corresponding lemma in Section 5.

Lemma 1 ([7]). *Let F be a function that takes $\{0, 1\}^n$ to $\{0, 1\}$. Suppose that given any $a, b \in$*

$\{0, 1\}^{n/2}$, one can construct gadget sequences $G(a)$ and $\overline{G}(b)$ of length L and L' , respectively, such that for an integer Y , for all $a, b \in \{0, 1\}^{n/2}$,

- if $F(a \odot b) = 1$, then $LCS(G(a), \overline{G}(b)) = Y$, and
- if $F(a \odot b) = 0$, then $LCS(G(a), \overline{G}(b)) \leq Y - 1$.

Then, one can construct two sequences A, B of length $2^{n/2} \text{poly}(L, L')$ such that for an integer E ,

- $LCS(A, B) = E$ if there exist $a, b \in \{0, 1\}^{n/2}$ such that $F(a \odot b) = 1$, and
- $LCS(A, B) \leq E - 1$ otherwise.

Armed with this lemma, we see that in order to prove our theorem, it suffices to create sequence gadgets G and \overline{G} of length $T^{O(\log W)}$ such that for some Y , $LCS(G(a), \overline{G}(b)) = Y$ if on input $a \odot b$, the starting state of the branching program reaches the accepting state, and $LCS(G(a), \overline{G}(b)) \leq Y - 1$ otherwise.

To construct $G(a)$ and $\overline{G}(b)$ we follow an inductive approach, mimicking Savitch's theorem [40]. Note that at this point the input is fixed, but we must implement $G(a)$ and $\overline{G}(b)$ independently. Let P be the given branching program of length T and width W , and assume for simplicity that $T = 2^t + 1$ for some $t \geq 0$. Since a and b are fixed, $G(a)$ and $\overline{G}(b)$ represent the corresponding subsets of edges of P , and the goal is to decide if there is a directed path from u_{start} to u_{acc} in the resulting graph. Such a path must go through some node in layer $2^{t-1} + 1$, and if we can guess which node then we can split P into two branching programs of half the length and evaluate each part recursively. We use a *reachability gadget* to implement this decomposition, and an LCS algorithm must then make the correct guess to find the longest common subsequence. The construction is thus recursive, and it roughly works as follows.

At the k -th level of the recursion we are given two nodes $u \in L_i$ and $v \in L_j$ with $j - i = 2^k$, and we want to decide if there is a directed path from u to v . We denote the sequence constructed in this case for a by $\text{RG}_k^{u \rightarrow v}(a)$ and for b by $\overline{\text{RG}}_k^{u \rightarrow v}(b)$. In particular $G(a) = \text{RG}_t^{u_{\text{start}} \rightarrow u_{\text{acc}}}(a)$ and $\overline{G}(b) = \overline{\text{RG}}_t^{u_{\text{start}} \rightarrow u_{\text{acc}}}(b)$. We define the sequences such that for some Y_k ,

- $LCS(\text{RG}_k^{u,v}(a), \overline{\text{RG}}_k^{u,v}(b)) = Y_k$ if on input $a \odot b$, one can reach v from u in 2^k steps, and
- $LCS(\text{RG}_k^{u,v}(a), \overline{\text{RG}}_k^{u,v}(b)) \leq Y_k - 1$ otherwise.

For $k = 0$, u and v are in neighboring layers, and they are connected if and only if there is an edge from u to v . Whether this is the case depends on the variable $x(i) = x_{f(i)}$, which either belongs to a or b . The sequence with no control over the edge is assigned a fixed symbol e , and the other sequence is assigned e if and only if the edge is present. It follows that $Y_0 = 1$. For $k > 0$, we inductively construct $\text{RG}_k^{u,v}(a)$ and $\overline{\text{RG}}_k^{u,v}(b)$ from the W choices $\text{RG}_{k-1}^{u,y}(a)$ and $\overline{\text{RG}}_{k-1}^{y,v}(b)$ where y is one of the W nodes in the layer of the branching program right in the middle between the layers of u and v .

Our construction ensures that the length of the gadgets for each k is $W^{O(k)}$, so that when we apply Lemma 1 we obtain sequences of length $2^{n/2} \cdot T^{O(\log W)}$, completing the proof.

Weighted LCS. To simplify the proof we will work with the following generalized version of LCS in which each letter in the alphabet can have a different weight. For two sequences P_1 and P_2 of length N over an alphabet Σ and a weight function $w : \Sigma \rightarrow [K]$, let X be the sequence that appears in both P_1, P_2 as a subsequence and maximizes the expression $w(X) = \sum_{i=1}^{|X|} w(X[i])$. We say that X is the *Weighted Longest Common Subsequence* (WLCS) of P_1, P_2 and write $\text{WLCS}(P_1, P_2) = w(X)$. The WLCS problem asks us to output $\text{WLCS}(P_1, P_2)$.

Note that a common subsequence X of two sequences P_1, P_2 can be thought of as an alignment or a matching $A = \{(a_i, b_i)\}_{i=1}^{|X|}$, where $a_i, b_i \in \mathbb{N}$ are indices, between the two sequences, so that for all $i \in [|X|] : P_1[a_i] = P_2[b_i]$, and $a_1 < \dots < a_{|X|}$ and $b_1 < \dots < b_{|X|}$. Clearly, the weight $\sum_{i=1}^{|X|} w(P_1[a_i]) = \sum_{i=1}^{|X|} w(P_2[b_i])$ of the matching A corresponds to the weighted length $w(X)$ of the common subsequence X .

Abboud *et al.* [7] showed a simple reduction from WLCS on length N sequences over an alphabet Σ with largest weight K to LCS on (unweighted) sequences of length $N \cdot K$ over the same alphabet. The reduction simply copies each symbol $\ell \in \Sigma$ in each of the sequences $w(\ell)$ times and then treats the sequences as unweighted. Abboud *et al.* showed that the optimal solution is preserved under this reduction.

For a sequence over a weighted alphabet Σ we define the *total length* of the sequence to be the sum of the weights of all symbols in this sequence. Note that this is the real length of the unweighted sequence that we obtain after applying the reduction from WLCS to LCS.

Reachability gadgets. The main component in our reduction are recursive constructions of two kinds of *reachability gadgets* $\text{RG}(a)$ and $\overline{\text{RG}}(b)$, with the following very useful property. For every pair of vectors $a \in A, b \in B$ and pair of nodes in the branching program P , u from layer i and v from layer j , such that $j - i = 2^k$ is a power of two, we have that:

- The LCS of the two sequences $\text{RG}_k^{u \rightarrow v}(a)$ and $\overline{\text{RG}}_k^{u \rightarrow v}(b)$ is equal to a certain fixed value Y_k that depends only on k if u can reach v with a path of the branching program that is induced by the assignment $a \odot b$, and the LCS is less than Y_k otherwise.
- The total length Z_k of any of these gadgets can be upper bounded by W^{ck} for some constant c .

We will now show how to construct such gadgets and upper bound their lengths.

Let $k \in \{0, \dots, \log_2 s\}$. We will inductively show how to construct, for any u, v , $\text{RG}_k^{u \rightarrow v}(a)$ and $\overline{\text{RG}}_k^{u \rightarrow v}(b)$ from $\text{RG}_{k-1}^{u \rightarrow y}(a)$ and $\overline{\text{RG}}_{k-1}^{y \rightarrow v}(b)$ for W nodes y .

Base Case: $k = 0$. Let u and v be two nodes lying in adjacent layers, i.e. u is in some layer i and v is in layer $i + 1$. In the base case, we check whether the layer i corresponds to an input variable $x(i)$ from X_1 or from X_2 . Assume that $x(i)$ is from X_1 . Let $\eta \in \{0, 1\}$ be the boolean value that a assigns to $x(i)$. Then check whether (u, v) is an edge in P and whether it is marked with η . If both conditions hold, we define $\text{RG}_0^{u \rightarrow v}(a) = e$ (for a letter e in the alphabet) and otherwise we define $\text{RG}_0^{u \rightarrow v}(a) = \1 (where $\$1$ is a letter that will never appear in the other sequence, so this is equivalent to defining this sequence to be the empty string). On the other hand, we unconditionally define $\overline{\text{RG}}_0^{u \rightarrow v}(b) = e$, because b is irrelevant for the current layer. In the symmetric case when $x(i)$ is from X_2 , set $\overline{\text{RG}}_0^{u \rightarrow v}(b) = e$ if (u, v) is an edge marked with the value assigned by b to $x(i)$. Otherwise, set $\overline{\text{RG}}_0^{u \rightarrow v}(b) = \2 , a symbol that does not appear in the other sequence. We set $w(e) = w(\$1) = w(\$2) = 1$.

Inductive step for $k > 0$. Now, let u be in layer i and v be in layer $j = i + 2^k$. We define $h = \frac{i+j}{2}$ to be the layer in the middle between i and j and note that $h - i = j - h = 2^{k-1}$. For each node $y = (h, z)$ for $z \in [W]$ in layer h we will add a gadget that enables the path from u to v to pass through this node y . We do this by recursively adding the two gadgets $\text{RG}_{k-1}^{u \rightarrow y}$ and $\overline{\text{RG}}_{k-1}^{y \rightarrow v}$. Layer k introduces $W + 3$ letters $f_k, g_k, \#_k$ and z_k for each $z \in [W]$.

$$\text{RG}_k^{u \rightarrow v}(a) = f_k^{2W} \left(\bigcup_{z \in [W]} g_k \underbrace{(z_k [\text{RG}_{k-1}^{u \rightarrow (h,z)}(a)] z_k [\text{RG}_{k-1}^{(h,z) \rightarrow v}(a)] z_k)}_{\text{The Core Gadget}} g_k \right) f_k^{2W} \#_k^{4W(W-1)}$$

The letter $\#_k$ will not appear in the other sequence and is completely unnecessary - we include it to make sure that both sequences have equal length and simplify the exposition. The padding of f_k and g_k is different in the other sequence, representing the partial assignments b to X_2 .

$$\overline{\text{RG}}_k^{u \rightarrow v}(b) = \bigcup_{x \in [W]} \left(f_k g_k^{2W} \underbrace{(z_k [\overline{\text{RG}}_{k-1}^{u \rightarrow (h,z)}(b)] z_k [\overline{\text{RG}}_{k-1}^{(h,z) \rightarrow v}(b)] z_k)}_{\text{The Core Gadget}} g_k^{2W} f_k \right).$$

The weights of the letters for layers $k > 0$. Our weights will guarantee that $w(f_k) = w(g_k) = w(\#_k)$. Assume that the total weight of any gadget $\text{RG}_{k-1}^{u \rightarrow v}(a)$ or $\overline{\text{RG}}_{k-1}^{u \rightarrow v}(b)$ is exactly some value Z_{k-1} . Then, the total weight of a gadget $\text{RG}_k^{u \rightarrow v}(a)$ or $\overline{\text{RG}}_k^{u \rightarrow v}(b)$ is fixed to $Z_k = W \cdot ((2W + 2) \cdot w(f_k) + 3w(z_k) + 2Z_{k-1})$, no matter what u, v, a, b are.

Let z represent a number in $[W]$. We define the remaining weights as:

$$w(z_k) = 2Z_{k-1}, \quad w(f_k) = w(g_k) = w(\#_k) = 9Z_{k-1}$$

Note that $w(f_k)$ is larger than the total weight of the “core gadget” which is $3w(z_k) + 2Z_{k-1} = 8Z_{k-1}$. This implies that $Z_k = Z_{k-1} \cdot W(9(4W + 2) + 8) = [W(36W + 26)]^k \leq W^{8k}$ assuming $W \geq 2$. For the base case we have $Z_0 = w(e) = 1$.

The LCS between two gadgets. We will compute the LCS by induction on $k = 0 \dots \log_2 T$. Let Y_k denote the LCS in the case that v is reachable from u , and let X_k denote the LCS in the complementary case. We will show that $Y_k = f(k, W) \leq W^{O(k)}$ and $X_k \leq Y_k - 1$.

In the base case, note that the LCS of $\text{RG}_0^{u \rightarrow v}(a)$ and $\overline{\text{RG}}_0^{u \rightarrow v}(b)$ is 1 if the edge $u \rightarrow v$ is consistent with the assignment $a \odot b$ while the LCS is 0 otherwise. Therefore, $Y_1 = 1$ and $X_1 = 0$.

Assume correctness for $k - 1$ and we will show it for k . We will first prove two claims.

Claim 1. *In an optimum WLCS, all letters f_k in $\overline{\text{RG}}_k^{u \rightarrow v}(b)$ are matched to f_k s in $\text{RG}_k^{u \rightarrow v}(a)$.*

Claim 2. *In an optimum WLCS, all letters g_k in $\text{RG}_k^{u \rightarrow v}(a)$ are matched to g_k s in $\overline{\text{RG}}_k^{u \rightarrow v}(b)$.*

We begin by proving Claim 1. Suppose that some number of f_k s in $\overline{\text{RG}}_k^{u \rightarrow v}(b)$ are not matched. Without loss of generality the first and last f_k are both matched. All other letters f_k appear in pairs one after the other. If one of the f_k is unmatched, then so is the other. This is because if the other was matched, it's matched either to the left or the right sequence of $2W$ f_k in $\text{RG}_k^{u \rightarrow v}(a)$, and since there are exactly $2W$ letters f_k in $\overline{\text{RG}}_k^{u \rightarrow v}(b)$, then one could also match the second f_k from

the pair and thus increasing the weight of the subsequence. Hence, we can assume that y pairs of f_k are unmatched for some y .

Now, if y pairs are unmatched, then at most $y + 1$ core gadgets from $\overline{\text{RG}}_k^{u \rightarrow v}(b)$ can be matched to $\text{RG}_k^{u \rightarrow v}(a)$. The potential gain in weight over the case when all f_k s in $\overline{\text{RG}}_k^{u \rightarrow v}(b)$ and all g_k s in $\text{RG}_k^{u \rightarrow v}(a)$ are matched is due to matching core gadgets:

$$(y + 1)8Z_{k-1} - 2y \cdot 9Z_{k-1} = (8 - 10y)Z_{k-1} < 0,$$

whenever $y \geq 1$. Thus, all f_k letters in $\overline{\text{RG}}_k^{u \rightarrow v}(b)$ are matched, and we've shown Claim 1.

Now we prove Claim 2. Since all f_k in $\overline{\text{RG}}_k^{u \rightarrow v}(b)$ are matched, there's a single inner gadget

$$g_k^{2W} (z_k [\overline{\text{RG}}_{k-1}^{u \rightarrow (h,z)}(b)] z_k [\overline{\text{RG}}_{k-1}^{(h,z) \rightarrow v}(b)] z_k) g_k^{2W}$$

for some z that can be matched for some z' to

$$\cup_{z' \in [W]} (g_k (z'_k [\text{RG}_{k-1}^{u \rightarrow (h,z')}(a)] z'_k [\text{RG}_{k-1}^{(h,z') \rightarrow v}(a)] z'_k) g_k).$$

Similar to the above, if some g_k from the second term is not matched, then there's a pair that is not matched. If y pairs are not matched, then the only potential gain is if $y + 1$ core gadgets get matched. The net gain is then again

$$(y + 1)8Z_{k-1} - 2y \cdot 9Z_{k-1} = (8 - 10y)Z_{k-1} < 0,$$

for any $y > 0$. Thus, all g_k s are matched, and this proves Claim 2.

Now, due to the above two claims, two core gadgets $(z_k [\overline{\text{RG}}_{k-1}^{u \rightarrow (h,z)}(b)] z_k [\overline{\text{RG}}_{k-1}^{(h,z) \rightarrow v}(b)] z_k)$ and $(z'_k [\text{RG}_{k-1}^{u \rightarrow (h,z')}(a)] z'_k [\text{RG}_{k-1}^{(h,z') \rightarrow v}(a)] z'_k)$ are aligned in some way by the optimum subsequence. If $z \neq z'$, then we lose $6Z_{k-1}$, potentially gaining only $2Z_{k-1}$ from matching the reachability gadgets for $k - 1$. This is a net loss, so we can assume that the optimum WLCS matches core gadgets for the same z . Now, the outer z_k will be matched since they don't interfere with the rest. If the middle z_k are not matched, then we lose $2Z_{k-1}$. We can gain at most $2Z_{k-1} - 1$ since the only way to gain $2Z_{k-1}$ is if both $\overline{\text{RG}}_{k-1}^{u \rightarrow (h,z)}(b)$ and $\overline{\text{RG}}_{k-1}^{(h,z) \rightarrow v}(b)$ are completely matched to their counterparts. However, then z_k would have the opportunity to be matched as well. Thus, we again have a net loss, so that $z = z'$ and all three z_k are matched. This means that $\overline{\text{RG}}_{k-1}^{u \rightarrow (h,z)}(b)$ is aligned with $\text{RG}_{k-1}^{u \rightarrow (h,z)}(a)$ and $\overline{\text{RG}}_{k-1}^{(h,z) \rightarrow v}(b)$ is aligned with $\text{RG}_{k-1}^{(h,z) \rightarrow v}(a)$ in some way.

This means that the LCS will be equal to $4w \cdot w(f_k) + 3w(z_k)$ plus the LCS of $\text{RG}_{k-1}^{u \rightarrow (h,z)}(a)$ and $\overline{\text{RG}}_{k-1}^{u \rightarrow (h,z)}(b)$ plus the LCS of $\text{RG}_{k-1}^{(h,z) \rightarrow v}(a)$ and $\overline{\text{RG}}_{k-1}^{(h,z) \rightarrow v}(b)$.

Here we use the key observation that u can reach v if and only if there is a node (h, z) for some $z \in [W]$ in the layer h that is between the layers of u and v such that both u can reach (h, z) and (h, z) can reach v . This implies that: The latter two contributions, by our inductive hypothesis, are $2 \cdot Y_{k-1}$ if there is a path (because we are free to pick the correct $z \in [W]$), and at most $Y_{k-1} + X_{k-1}$ if there isn't one (no matter what z we pick, this will be the contribution). We get:

$$Y_k = 4W \cdot w(f_k) + 3w(z_k) + 2Y_{k-1} = 36W \cdot Z_{k-1} + 6Z_{k-1} + 2Y_{k-1} = 6(6W + 1)Z_{k-1} + 2Y_{k-1}.$$

and:

$$X_k \leq 4W \cdot w(f_k) + 3w(z_k) + Y_{k-1} + X_{k-1} \leq Y_k - (Y_{k-1} - X_{k-1}) \leq Y_k - 1,$$

which proves the inductive step.

The final vector gadgets. Assume that the branching program has a fixed start node $u_{\text{start}} = (1, 1)$ in the first layer and a single fixed “accept” node $u_{\text{acc}} = (T, 1)$ in the last layer. The branching program P accepts on input $a \odot b$ if and only if the path induced by the assignment $a \odot b$ that starts at u_{start} reaches u_{acc} . We will map each vector $a \in A$ in the first list to the vector gadget $VG'(a) = \text{RG}_k^{u_{\text{start}} \rightarrow u_{\text{acc}}}(a)$ where $k = \log_2 T$ which we assume without loss of generality is an integer, and we map each vector $b \in B$ in the second list to the vector gadget $VG'(b) = \overline{\text{RG}}_k^{u_{\text{start}} \rightarrow u_{\text{acc}}}(b)$. We define the constant $Y = Y_k$, for $k = \log_2 T$.

By the discussion above we have that the LCS of $VG'(a)$ and $VG'(b)$ is equal to Y if our branching program P accepts on input $a \odot b$, and the LCS is at most $Y - 1$ otherwise. Moreover, we have that the total length of these gadgets can be upper bounded by $W^{8 \log_2 T} = T^{8 \log W}$.

We finish the proof of the theorem by applying Lemma 1 from the beginning of the section.

3.1 Reduction to k -LCS

A nice property of the reduction for LCS in Theorem 3 is that it can be extended to a reduction to K -LCS, we give a detailed sketch of the proof below.

Theorem 4. *Given a branching program of length T , width W and n inputs, one can produce for any $K \geq 2$, in time $2^{n/K} K^{O(K)} T^{O(\log W)}$ an instance of K -LCS of alphabet size $O((K + W) \log T)$, where the sequences have length $2^{n/K} K^{O(K)} T^{O(\log W)}$.*

As with LCS, we begin with a Lemma from prior work [7] that allows us to select a K -tuple of partial assignments that together form an accepting input to the branching program.

Lemma 2 (ABV'15). *Let F be a function that takes $\{0, 1\}^n$ to $\{0, 1\}$. Suppose that given any $\alpha \in \{0, 1\}^{n/k}$, one can construct k gadget sequences $G_i(\alpha)$ for $i \in [k]$ of length at most D , such that for an integer Y , for all $\alpha_1, \alpha_2, \dots, \alpha_k \in \{0, 1\}^{n/k}$,*

- *if $F(\odot_{i=1}^k \alpha_i) = 1$, then $\text{LCS}(\{G_1(\alpha_1), \dots, G_k(\alpha_k)\}) = Y$, and*
- *if $F(\odot_{i=1}^k \alpha_i) = 0$, then $\text{LCS}(\{G_1(\alpha_1), \dots, G_k(\alpha_k)\}) \leq Y - 1$.*

Then, one can construct k sequences A_1, \dots, A_k each of length $\leq 2^{n/k} k^{O(k)} D^{O(1)}$ such that for an integer E ,

- *$\text{LCS}(\{A_1, \dots, A_k\}) = E$ if there exist $\alpha_1, \alpha_2, \dots, \alpha_k \in \{0, 1\}^{n/k}$ such that $F(\odot_{i=1}^k \alpha_i) = 1$, and*
- *$\text{LCS}(\{A_1, \dots, A_k\}) \leq E - 1$ otherwise.*

Armed with this Lemma, it suffices to extend our reachability gadget construction for checking consistency of K partial assignments. We will be able to do this without increasing the total weight of the gadgets. Just as with the LCS proof, we work with weighted K -LCS, which from [7] we know how to reduce back down to K -LCS where the length of the sequences is the total weight. We provide a sketch.

We will reduce to K -LCS where $K \geq 2$ is any integer. We split the input variables x_1, \dots, x_n to the branching program P in K sets, $X_i = \{x_{(i-1)n/K+1}, \dots, x_{in/K}\}$ for $i \in [K]$.

The reachability gadget for the i -th partial assignment α_i (to the variables in X_i) and the i th sequence will be denoted by $\text{RG}_{i,k}^{u \rightarrow v}(\alpha_i)$.

Base Case: $k = 0$. Let u and v be two nodes lying in adjacent layers, i.e. u is in some layer i and v is in layer $i + 1$. We describe the reachability gadget $\text{RG}_{j,0}^{u \rightarrow v}(\alpha_j)$. For each j we will have a symbol $\$j$ that does not appear in any sequence other than the j th. We also have a letter e that can appear in any sequence.

Fix j . As before, we check whether the layer i corresponds to an input variable $x(i)$ from X_j . Assume first that $x(i)$ is from X_j . Let $\eta \in \{0, 1\}$ be the boolean value that α_j assigns to $x(i)$. Then check whether (u, v) is an edge in P and whether it is marked with η . If both conditions hold, we define $\text{RG}_{j,0}^{u \rightarrow v}(\alpha_j) = e$ and otherwise we define $\text{RG}_{j,0}^{u \rightarrow v}(\alpha_j) = \j . If $x(i)$ is not from X_j , we set $\text{RG}_{j,0}^{u \rightarrow v}(\alpha_j) = e$. We set $w(e) = w(\$j) = 1$ for all $j \in [K]$.

Inductive step for $k > 0$. Now, let u be in layer i and v be in layer $i + 2^k$. We define $h = \frac{i + (i + 2^k)}{2} = i + 2^{k-1}$ to be the layer in the middle between i and $i + 2^k$. For each node $y = (h, z)$ for $z \in [W]$ in layer h we will add a gadget that enables the path from u to v to pass through this node y .

Layer k introduces $W + K + 1$ letters $f_k, g_k, \#_{j,k}$ and z_k for each $z \in [W]$.

Let $j \leq K - 1$. We define

$$\text{RG}_{j,k}^{u \rightarrow v}(\alpha_j) = f_k^{2W} \left(\bigcup_{z \in [W]} g_k \underbrace{(z_k [\text{RG}_{j,k-1}^{u \rightarrow (h,z)}(\alpha_j)] z_k [\text{RG}_{j,k-1}^{(h,z) \rightarrow v}(\alpha_j)] z_k)}_{\text{The Core Gadget}} g_k \right) f_k^{2W} \#_{j,k}^{4W(W-1)}$$

The letter $\#_{j,k}$ will not appear in the other sequences and is completely unnecessary - we include it to make sure that both sequences have equal length and simplify the exposition. The padding of f_k and g_k is different in the K th sequence, representing the partial assignments to X_K .

$$\text{RG}_{K,k}^{u \rightarrow v}(\alpha_K) = \bigcup_{x \in [W]} \left(f_k g_k^{2W} \underbrace{(z_k [\text{RG}_{K,k-1}^{u \rightarrow (h,z)}(\alpha_K)] z_k [\text{RG}_{K,k-1}^{(h,z) \rightarrow v}(\alpha_K)] z_k)}_{\text{The Core Gadget}} g_k^{2W} f_k \right).$$

As before, assume that the total weight of any gadget $\text{RG}_{j,k-1}^{u \rightarrow v}(\alpha_j)$ is exactly some value Z_{k-1} , and that we set $w(f_k) = w(g_k) = w(\#_{j,k})$ for all j . Then, the total weight of a gadget $\text{RG}_{j,k}^{u \rightarrow v}(\alpha_j)$ is fixed to $Z_k = W \cdot ((2W + 2) \cdot w(f_k) + 3w(z_k) + 2Z_{k-1})$, no matter what u, v, α_j are.

As before, pick $w(f_k) = w(g_k) = w(\#_{j,k}) = 9Z_{k-1}$ and $w(z_k) = 2Z_{k-1}$ for all $z \in [W]$. The total weight of the “core gadget” is again $3w(z_k) + 2Z_{k-1} = 8Z_{k-1}$, and $Z_k = Z_{k-1} \cdot W(9(4W + 2) + 8) = [W(36W + 26)]^k \leq W^{8k}$ assuming $W \geq 2$. For the base case we have $Z_0 = w(e) = 1$.

As before, if any f_k from $\text{RG}_{K,k}^{u \rightarrow v}(\alpha_K)$ is not matched, then again a pair of adjacent ones isn't matched. If y pairs are not matched, then at most $y + 1$ inner gadgets are, and the net gain is negative as before. Hence, in an optimum alignment, all f_k from $\text{RG}_{K,k}^{u \rightarrow v}(\alpha_K)$ are matched. This selects some z so that

$$g_k^{2W} (z_k [\text{RG}_{K,k-1}^{u \rightarrow (h,z)}(b)] z_k [\text{RG}_{K,k-1}^{(h,z) \rightarrow v}(b)] z_k) g_k^{2W}$$

is aligned with the inner gadgets

$$\bigcup_{z^j \in [W]} (g_k (z_k^j [\text{RG}_{j,k-1}^{u \rightarrow (h,z^j)}(a)] z_k^j [\text{RG}_{j,k-1}^{(h,z^j) \rightarrow v}(a)] z_k^j) g_k),$$

for the rest of the sequences $j \leq K - 1$. Now, as before, all g_k letters from each of the inner gadgets above must be matched, as otherwise we would lose too much weight. Hence the alignment is between

$$\{z_k^j [\text{RG}_{j,k-1}^{u \rightarrow (h, z^j)}(a)] z_k^j [\text{RG}_{k-1}^{(h, z^j) \rightarrow v}(a)] z_k^j\}_{j \in [K]},$$

for potentially different z^j s. Because all g_k s are matched, we must have $z^j = z^i$ for all $i, j < K$, but potentially it could be that $z^1 \neq z^K$. However, as $w(z_k)$ is large, again, the alignment needs to be for the same z , and all three z_k s in the core gadgets must be matched. This completes the proof sketch.

4 Sequence Problems with Alignment Gadgets

Bringmann and Künnemann [22] introduced a framework for showing SETH lower bounds for problems that compute a similarity measure of two sequences of, e.g., bits, symbols, or points. They showed that any sequence-problem that implements so-called *alignment gadgets* cannot be solved in truly subquadratic time under SETH. We show that alignment gadgets are actually much stronger. So strong, in fact, that they can simulate nondeterministic branching programs. It follows that Edit-Distance, Longest Common Subsequence (LCS), Dynamic Time Warping Distance, and every other problem that implements alignment gadgets can solve SAT on nondeterministic branching programs. This proves our main result, Theorem 2. We start here with a sketch of the proof, and then provide the details in Section 5.

A similarity measure δ is a function that measures the similarity, e.g., Edit-Distance or LCS, of two given sequences A and B . Suppose $|A| \geq |B|$. A *structured alignment* maps B to a consecutive subsequence A' of A , and the cost of the alignment is $\delta(A', B)$. Note that the alignment does not correspond to, e.g., a common subsequence. Instead the alignment restricts the similarity measure to only work with A' and B . An alignment gadget is a gadget that takes two collections of sequences and combines each collection into a single sequence such that the minimum cost of a structured alignment closely approximates (within an additive constant) the similarity measure of the two constructed sequences. An alignment gadget can thus be interpreted as a way of forcing the similarity measure to respect the structure of the two collections of sequences. Alignment gadgets are then combined recursively in order to obtain a reduction. We next sketch how this is done for branching programs.

Let P be a given branching program on n boolean inputs, and let F be the corresponding function. To prove a reduction from BP-SAT to a sequence-problem with alignment gadgets we again go through the satisfying pair problem on branching programs (see Section 2). Let therefore $X_1 = \{x_1, \dots, x_{n/2}\}$ and $X_2 = \{x_{n/2+1}, \dots, x_n\}$ be the first and last half of the inputs to the branching program, respectively. We must decide whether there exist $a, b \in \{0, 1\}^{n/2}$ such that $F(a \odot b) = 1$, where $a \odot b$ is the concatenation of a and b .

Our reduction uses alignment gadgets to construct for each $a \in \{0, 1\}^{n/2}$ a sequence $G(a)$, and for each $b \in \{0, 1\}^{n/2}$ another sequence $\overline{G}(b)$. These sequences are constructed such that their similarity measure is $\delta(G(a), \overline{G}(b)) = Y$, for some integer Y , if $F(a \odot b) = 1$; and such that $\delta(G(a), \overline{G}(b)) \geq Y + \rho$ otherwise, where $\rho > 0$ is the same for all a and b . In previous reductions from OV the construction of $G(a)$ and $\overline{G}(b)$ was nearly trivial, but as in Section 3 it is now the main challenge when proving our reduction.

Once we have constructed $G(a)$ and $\overline{G}(b)$ for all $a, b \in \{0, 1\}^{n/2}$, we combine them into two sequences that will be the output of the reduction. This step is almost identical to a corresponding

step in the reduction by Backurs and Indyk [16] from orthogonal vectors to Edit-Distance, and later by Abboud *et al.* [7] to LCS, and by Bringmann and Künnemann [22] to sequence-problems with alignment gadgets. If there exist $a_i, b_j \in \{0, 1\}^{n/2}$ with $F(a_i \odot b_j) = 1$, then the goal is to make the structured alignment match $G(a_i)$ and $\overline{G}(b_j)$. It is therefore tempting to apply the alignment gadget to $A = (G(a_1), \dots, G(a_N), G(a_1), \dots, G(a_N))$ and $B = (\overline{G}(b_1), \dots, \overline{G}(b_N))$, where $N = 2^{n/2}$, since we can then freely map B to a consecutive subsequence of A such that $\overline{G}(b_j)$ maps to $G(a_i)$. The contribution to the similarity measure from this pair would then be Y , but unfortunately we have no control over the rest of the sequence. To finish the proof we therefore need one more idea: the *normalized vector gadget*. We put a dummy sequence next to every subsequence in A , such that sequences in B always have the option of mapping to dummy sequences, thereby contributing $Y + \rho$ to the similarity measure. (The alignment gadgets framework allows gadgets of different *types* that are implicitly padded to some specified length. This is used to handle the technicality that, e.g., the length of subsequences in B no longer correctly match those in A .) We finally get that if F evaluates to 1 for a pair of inputs then $\delta(A, B) \leq N(Y + \rho) - \rho$, and otherwise $\delta(A, B) = N(Y + \rho)$. This completes the reduction.

Simulating branching programs with alignment gadgets. To construct $G(a)$ and $\overline{G}(b)$ we essentially implement the proof of Savitch's theorem with alignment gadgets. The construction is almost the same as the one given in Section 3. We again note that the input is fixed at this point, and that we must implement $G(a)$ and $\overline{G}(b)$ independently. Let P be the given branching program of length T and width W , and assume for simplicity that $T = 2^t + 1$ for some $t \geq 0$. Since a and b are fixed, $G(a)$ and $\overline{G}(b)$ represent the corresponding subsets of edges of P , and the goal is to decide if there is a directed path from u_{start} to u_{acc} in the resulting graph. Such a path must go through some node in layer $2^{t-1} + 1$, and if we can guess which node then we can split P into two branching programs of half the size and evaluate each part recursively. We use an alignment gadget to implement this decomposition, and we use the similarity measure to make the correct guess. The construction is thus recursive, and it works as follows.

At the k -th level of the recursion we are given two nodes $u \in L_i$ and $v \in L_j$ with $j - i = 2^k$, and we want to decide if there is a directed path from u to v . We denote the sequence constructed in this case for a by $\text{RG}_X^{k, u \rightarrow v}(a)$ and for b by $\text{RG}_Y^{k, u \rightarrow v}(b)$. In particular $G(a) = \text{RG}_X^{t, u_{\text{start}} \rightarrow u_{\text{acc}}}(a)$ and $\overline{G}(b) = \text{RG}_Y^{t, u_{\text{start}} \rightarrow u_{\text{acc}}}(b)$. (Note that the notation differs slightly from the notation used in Section 3. The parameters X and Y are used to determine the types of the sequences, and thereby their implicit length, but we ignore that aspect here.) For $k = 0$, u and v are in neighboring layers, and they are connected if and only if there is an edge from u to v . Whether this is the case depends on the variable $x(i) = x_{f(i)}$, which either belongs to a or b . The sequence with no control over the edge is assigned a 1, and the other sequence is assigned 1 if the edge is present and 0 otherwise. The contribution to the similarity measure is thus 0 if the edge is present and 1 otherwise. (Such unit sequences are called *coordinate values* in the framework, and it requires a proof to show that the problem in question supports them. Bringmann and Künnemann [22] provided the relevant proofs for our purposes.)

For $k > 0$, we use an alignment gadget to pick the node that the path from u to v passes through. Let $\ell = (i + j)/2$. Ideally we would like to apply the alignment gadget to sequences of

the form

$$\begin{aligned} A &= \bigcup_{z \in L_\ell} (\text{RG}_X^{k-1, u \rightarrow z}(a), \text{RG}_X^{k-1, z \rightarrow v}(a)) \\ B &= \bigcup_{z \in L_\ell} (\text{RG}_Y^{k-1, u \rightarrow z}(b), \text{RG}_Y^{k-1, z \rightarrow v}(b)) \end{aligned}$$

and then let the alignment match $(\text{RG}_X^{k-1, u \rightarrow z}(a), \text{RG}_X^{k-1, z \rightarrow v}(a))$ and $(\text{RG}_Y^{k-1, u \rightarrow z}(b), \text{RG}_Y^{k-1, z \rightarrow v}(b))$ for some $z \in L_\ell$. There are, however, two obstacles that we must overcome to make this work. First, the alignment may overlap sequences in undesired ways. In Section 3 we inserted special symbols of large weight between the recursively defined sequences, such that these symbols had to be matched correctly. We now introduce *index gadgets* that serve a similar function. Second, and more importantly, we must rule out that a different z is picked for A and B . In Section 3 this issue was handled by encapsulating the sequences with other, longer, trivial sequences (heavy symbols) to get the right behavior. We again use index gadgets to get this effect. (We also use an *OR gadget* that is composed of two alignment gadgets to put things together.)

Recall that W is the width. Although it may look like a sequence constructed at level k is $O(W)$ times longer than a sequence constructed at level $k - 1$, the extensive padding actually causes a blowup of a factor $O(W^2)$. The final sequences $G(a)$ and $\overline{G}(b)$ therefore have length $W^{O(\log T)}$. Correctness of the reduction follows from the fact that the similarity measure is lower for sequences that correspond to connected vertices. We have therefore arrived at the following technical theorem, which implies Theorem 2. The conditions for the theorem were proved by Bringmann and Künnemann [22] for, e.g., Edit-Distance, LCS, and Dynamic Time Warping.

Theorem 5. *Let δ be a similarity measure over sequences in \mathcal{I} admitting an alignment gadget of size $f(n) = cn$ and coordinate values, and consider the problem of computing $\delta(x, y)$ for sequences $x, y \in \mathcal{I}$ of length N . SAT of nondeterministic Branching Programs over n variables of width W and length T can be reduced to an instance of this problem on sequences where $N = O(n \cdot T^{\log_2(12W^2c^3)} \cdot c^2 \cdot \log_2 W) = T^{O(\log W)} \cdot n$.*

4.1 Definitions

We start by defining the building blocks from which we will implement our reduction. We will prove a generic reduction from BP SAT to a generic problem of computing a similarity measure δ of two sequences, for any δ that has a certain property. This property will be the ability to implement *alignment gadgets* which we define below. The advantage of this generic proof over a direct proof is that we can reduce the amount of case-analysis that is required in the proofs: many steps in the reduction will require similar functionalities, and this framework allows us to only prove this functionality once. We will borrow the notation of [22] and reintroduce their *alignment gadgets*.

Let $\delta : \mathcal{I} \times \mathcal{I} \rightarrow \mathbb{N}_0$ be a similarity measure for a pair of inputs from \mathcal{I} . As a running example it is convenient to think of \mathcal{I} as being all binary sequences of length n and let δ be the Edit Distance between two sequences. For a sequence $x \in \mathcal{I}$ we define its *type* to be a tuple encoding its length and the sum of its entries, i.e. $\text{type}(x) := (|x|, \sum_i x[i])$. For example, if x is a binary sequence, its type encodes its length $|x|$ and the number of ones in x . For a type $t \in \mathbb{N} \times \mathbb{N}_0$ we let $\mathcal{I}_t := \{x \in \mathcal{I} \mid \text{type}(x) = t\}$ be the set of all inputs of type t . We remark that the exact definition of *type* will not be crucial to the framework and for different measures it might be convenient to use a different definition.

Alignments Let $n \geq m$. A (partial) alignment is a set $A = \{(i_1, j_1), \dots, (i_k, j_k)\}$ with $0 \leq k \leq m$ such that $1 \leq i_1 < \dots < i_k \leq n$ and $1 \leq j_1 < \dots < j_k \leq m$. We say that a pair $(i, j) \in A$ are *aligned*. Let $\mathcal{A}_{n,m}$ be the set of all partial alignments with respect to n, m . A partial alignment of the form $\{(\Delta+1, 1), \dots, (\Delta+m, m)\}$ for some $0 \leq \Delta \leq n-m$ will be called a *structured alignment*, and we will denote the set of all such alignments by $\mathcal{S}_{n,m} \subseteq \mathcal{A}_{n,m}$.

Consider any sequences $x_1, \dots, x_n \in \mathcal{I}$ and $y_1, \dots, y_m \in \mathcal{I}$. Let $Q = \max_{i,j} \delta(x_i, y_j)$ be the maximum distance between a pair of sequences under our measure δ . We define the *cost* of an alignment $A \in \mathcal{A}_{n,m}$ (with respect to our sequences) by

$$\text{cost}(A) := \sum_{(i,j) \in A} \delta(x_i, y_j) + (m - |A|) \cdot Q$$

so that any $j \in [m]$ that is not aligned in A will incur the maximal cost Q .

Alignment Gadget Intuitively, an alignment gadget is a way to combine sequences $x_1, \dots, x_n \in \mathcal{I}$ into one sequence $X = \text{GA}(x_1, \dots, x_n)$, and at the same time, to combine sequences $y_1, \dots, y_m \in \mathcal{I}$ into one sequence $Y = \text{GA}(y_1, \dots, y_m)$, so that: $\delta(X, Y)$ is essentially equal to $\delta(A)$ for the optimal *structured* alignment $A \in \mathcal{S}_{n,m}$ with respect to our sequences. We will be interested in showing and deriving consequences of the existence of efficient implementations of such gadgets for a similarity measure δ . The formal definition introduces additional technicalities which simplify the proofs of existence of these gadgets considerably.

Definition 2 (Alignment Gadgets). *The similarity measure δ admits an alignment gadget of size $f(n)$, if the following conditions hold: Given instances $x_1, \dots, x_n \in \mathcal{I}_{\tau_X}, y_1, \dots, y_m \in \mathcal{I}_{\tau_Y}$ with $m \leq n$ and types $\tau_X = (\ell_X, s_X), \tau_Y = (\ell_Y, s_Y)$, we can construct new instances $x = \text{GA}_X^{m, \tau_Y}(x_1, \dots, x_n)$ and $y = \text{GA}_Y^{n, \tau_X}(y_1, \dots, y_m)$ and $C \in \mathbb{Z}$ such that:*

- $\min_{A \in \mathcal{A}_{n,m}} \text{cost}(A) \leq \delta(x, y) - C \leq \min_{A \in \mathcal{S}_{n,m}} \text{cost}(A)$.
- *type*(x) and *type*(y) only depend on n, m, τ_X , and τ_Y .
- $|x|, |y| \leq f(n) \cdot (\ell_X + \ell_Y)$.
- *This construction runs in $O(n(\ell_X + \ell_Y))$ time.*

The second ingredient that is required in order to implement a reduction from SAT problems to computing similarity measures are *coordinate values* which are atomic sequences of constant length that we will combine in various ways with the alignment gadgets.

Definition 3 (Coordinate Values). *The similarity measure δ admits coordinate values if there are instances $\mathbf{1}_X, \mathbf{0}_X, \mathbf{1}_Y, \mathbf{0}_Y \in \mathcal{I}$ such that:*

$$\delta(\mathbf{1}_X, \mathbf{1}_Y) > \delta(\mathbf{0}_X, \mathbf{1}_Y) = \delta(\mathbf{1}_X, \mathbf{0}_Y) = \delta(\mathbf{0}_X, \mathbf{0}_Y)$$

and *type*($\mathbf{1}_X$) = *type*($\mathbf{0}_X$) and *type*($\mathbf{1}_Y$) = *type*($\mathbf{0}_Y$).

4.2 OR Gadgets

In our reduction from BP SAT, it will be convenient to work with an OR gadget, besides an alignment gadget. Next, we define this gadget and then prove that any similarity measure δ that can implement alignment gadgets will also be able to implement OR gadgets (with some loss of efficiency).

Definition 4 (OR Gadgets). *The similarity measure δ admits OR gadgets of size $f(n)$, if the following conditions hold: Given instances $x_1, \dots, x_n \in \mathcal{I}_{\tau_X}, y_1, \dots, y_n \in \mathcal{I}_{\tau_Y}$ and types $\tau_X = (\ell_X, s_X), \tau_Y = (\ell_Y, s_Y)$, we can construct new instances*

$$x = \text{OR}_X^{\tau_Y}(x_1, \dots, x_n)$$

$$y = \text{OR}_Y^{\tau_X}(y_1, \dots, y_n)$$

and $C \in \mathbb{Z}$ such that:

- $\delta(x, y) = C + \min_{i,j \in [n]} \delta(x_i, y_j)$
- $\text{type}(x)$ and $\text{type}(y)$ only depend on n, τ_X , and τ_Y .
- $|x|, |y| \leq f(n) \cdot (\ell_X + \ell_Y)$.
- This construction runs in $O(n(\ell_X + \ell_Y))$ time.

By combining two alignment gadgets in a careful way, we obtain an OR gadget. Note that there is a quadratic blow up in the size of the gadgets in the following lemma. For this reason, we will only use OR gadgets when working with a small number of sequences.

Lemma 3. *Any similarity measure δ that admits alignment gadgets of size $f(n) = c \cdot n$, also admits OR gadgets of size $f'(n) = c^2 n^2$.*

Proof. Given instances $x_1, \dots, x_n \in \mathcal{I}_{\tau_X}, y_1, \dots, y_n \in \mathcal{I}_{\tau_Y}$ and types $\tau_X = (\ell_X, s_X), \tau_Y = (\ell_Y, s_Y)$, we will construct OR gadgets as follows. First, construct $x' := \text{GA}_X^{\tau_Y, 1}(x_1, \dots, x_n)$ and for all $j \in [n]$ construct $y'_j := \text{GA}_Y^{\tau_X, n}(y_j)$. Let $t'_X = \text{type}(x')$ and $t'_Y = \text{type}(y'_j)$ (and note that it is independent of j). Then, our final gadgets are $x := \text{GA}_X^{t'_Y, n}(x')$ and $y := \text{GA}_Y^{t'_X, 1}(y'_1, \dots, y'_n)$.

First, note that $\text{type}(x)$ and $\text{type}(y)$ only depend on n, τ_X and τ_Y . Then, let us bound the lengths of x, y . We know that $|x'|, |y'_j| \leq c \cdot n \cdot (\ell_X + \ell_Y)$, and therefore $|x|, |y| \leq c \cdot n \cdot (|x'| + |y'_j|) \leq c^2 \cdot n^2 \cdot (\ell_X + \ell_Y)$.

Finally, we prove the correctness. By definition of alignment gadgets, for all $j \in [n]$ we have $\delta(x', y'_j) = C + \min_{i \in [n]} \delta(x_i, y_j)$, for some fixed integer C . Moreover, for some integer C' , we have that $\delta(x, y) = C' + \min_{j \in [n]} \delta(x', y'_j)$ which is equal to $C' + C + \min_{j \in [n]} \min_{i \in [n]} \delta(x_i, y_j)$. \square

4.3 Similarity Measures with Alignment Gadgets

In their paper, Bringmann and Künnemann construct alignment gadgets for a few fundamental similarity measures. Combining these gadgets with our main theorem implies significantly stronger lower bounds for computing these measures. We list these measures and the corresponding sizes of alignments gadgets below.

Edit Distance. The Edit Distance between two sequences x, y is the minimum number of insertions, deletions, and substitutions that is required to transform one sequence to the other. The most basic case is when the sequences are binary, that is the set of instances \mathcal{I} is the set of binary sequences $\{0, 1\}^n$.

Lemma 4 ([22]). *There is a constant $c \leq 10^3$ such that Edit Distance similarity measure over binary sequences admits an alignment gadget of size $f(n) \leq c \cdot n$.*

It is likely that smaller alignment gadgets can be obtain if the alphabet size is larger.

Longest Common Subsequence. In Section 3 we showed a direct reduction from BP-SAT to LCS. We will use the alignment gadgets framework in order to obtain the same reduction but to sequences over *binary* inputs. That is, the surprising expressibility of LCS that is exhibited by our proofs is already present when we only have two distinct letters to match.

Lemma 5 ([22]). *There is a constant $c \leq 10^3$ such that Longest Common Subsequence similarity measure over binary sequences admits an alignment gadget of size $f(n) \leq c \cdot n$.*

Dynamic Time Warping Distance. The DTWD over curves in various metric spaces is of great practical interest. We are able to show lower bounds even in the special case of one-dimensional curves. Let $x, y \in \mathbb{Z}^n$ be two sequences of n integers. The DTWD $\delta_{DTWD}(x, y)$ of the two curves is the minimum cost of a joint traversal of both curves. A *traversal* of two curves is a process that places a marker at the beginning of each curve and during each step one or both markers are moved forward one point, until the end of both curves is reached. Each step aligns two points, one from each curve. The cost of a traversal is the sum of distances between all aligned points. In our case, the distance is simply the absolute value of the difference between the aligned integers.

Lemma 6 ([22]). *There is a constant $c \leq 10$ such that Dynamic Time Warping Distance similarity measure over one dimensional curves admits an alignment gadget of size $f(n) \leq c \cdot n$.*

5 The Full Reduction

We are now ready to prove our main theorem, from which Theorem 2 follows, by the Lemmas in the previous section.

Theorem 6 (Main). *Let δ be a similarity measure over sequences in \mathcal{I} admitting an alignment gadget of size $f(n) = cn$ and coordinate values, and consider the problem of computing $\delta(x, y)$ for sequences $x, y \in \mathcal{I}$ of length N . SAT of nondeterministic Branching Programs over n variables of width W and length T can be reduced to an instance of this problem on sequences where $N = O(n \cdot T^{\log_2(12W^2c^3)} \cdot c^2 \cdot \log_2 W) = T^{O(\log W)} \cdot n$.*

Proof. Let δ be a similarity measure that admits alignment gadgets and coordinate values. We will construct several other gadgets with certain properties from these primitives. Given an instance of SAT on BPs of on n variables width W and length T we will construct and combine our gadgets in a certain way into two sequences x, y such that $\delta(x, y)$ will determine whether our instance is “yes” or “no”.

Let A and B both be the set of all binary vectors of length $n/2$. The goal of the reduction is to find a pair $a \in A, b \in B$ such that our given branching program is satisfied by the assignment in which the first half of the variables are assigned according to a while the second half are assigned according to b .

We will use the parameters $w := \log_2 W$, and $t = \log_2 T$, and assume these are integers.

Reachability gadgets. The main component in our reduction are recursive constructions of two kinds of *reachability gadgets* $\text{RG}_X(a)$ and $\text{RG}_Y(b)$, with the following very useful property. For every pair of vectors $a \in A, b \in B$ and pair of nodes in the branching program $u \in L_i$ from layer i and $v \in L_j$ from layer j , such that $j - i = 2^{k-1}$ is a power of two, we have that:

- the δ -distance between the two sequences $\text{RG}_X^{k,u \rightarrow v}(a)$ and $\text{RG}_Y^{k,u \rightarrow v}(b)$ is equal to a certain fixed value ρ_k that depends only on k if the path of the branching starting at u and induced by the assignment (a, b) reaches v , and the δ -distance is greater than ρ_k otherwise.
- the total length of these gadgets can be upper bounded by $\ell_k \leq O(W^2 c^3)^k$.

We will now show how to construct such gadgets and prove that they satisfy the above properties. Then, we will use these gadgets in order to check whether a pair (a, b) makes the accept node u_{acc} reachable from the start node u_{start} .

Base Case: $k = 1$. We start by defining the gadgets $\text{RG}^{1,u \rightarrow v}$ for the case that $k = 1$ and our two vertices u, v are in consecutive layers i and $i + 1$ for some $i \in [s]$. To do this, we consider the variable x_j that the layer i in our branching program is labelled with, and check whether that variable appears in the vectors in A or in B – that is, we check whether $j \leq d/2$ or $j > d/2$ where d is the number of variables in our branching program.

In the first case, the vector a is “responsible” for verifying the consistency of the edge $u \rightarrow v$, and we define the gadgets as follows: we set

$$\text{RG}_X^{1,u \rightarrow v}(a) := EG_X(0) := \text{GA}_X^{2w+2, \tau_Y}(\mathbf{0}_X, \dots, \mathbf{0}_X, \mathbf{1}_X, \mathbf{0}_X)$$

if the edge $u \rightarrow v$ is labelled with the boolean value $\eta \in \{0, 1\}$ which is the same as the boolean value that a assigns to the variable x_j , and otherwise we set

$$\text{RG}_X^{1,u \rightarrow v}(a) = EG_X(1) := \text{GA}_X^{2w+2, \tau_Y}(\mathbf{1}_X, \dots, \mathbf{1}_X, \mathbf{1}_X, \mathbf{0}_X).$$

Note that inconsistency can either be caused by the edge not existing in the branching program, or by the vector assigning a different value to the corresponding variable. On the other hand, since the vector b is not “responsible” for the edge $u \rightarrow v$, we unconditionally set

$$\text{RG}_Y^{1,u \rightarrow v}(b) := EG_Y(1) := \text{GA}_Y^{2w+2, \tau_X}(\mathbf{1}_Y, \dots, \mathbf{1}_Y, \mathbf{0}_Y, \mathbf{1}_Y).$$

The reason for defining these gadgets as a concatenation of $2w + 2$ value gadgets instead of one is purely technical: we want to ensure that these gadgets have the same type as our “index gadgets” which we will define shortly. Also, the last two coordinates are supposed to distinguish between our “reachability gadgets” and our “index gadgets”, so that matching an index gadget with a reachability gadget will incur a loss due to these coordinates.

In the second case, the vector b is “responsible” for verifying the consistency of the edge $u \rightarrow v$, and we define the gadgets in a symmetric way: The gadget $\text{RG}_X^{1,u \rightarrow v}(a)$ will be unconditionally set to $EG_X(1)$. While we set

$$\text{RG}_Y^{1,u \rightarrow v}(b) := EG_Y(0) := \text{GA}_Y^{2w+2, \tau_X}(\mathbf{0}_Y, \dots, \mathbf{0}_Y, \mathbf{0}_Y, \mathbf{1}_Y)$$

if the edge $u \rightarrow v$ is labelled with the boolean value $\eta \in \{0, 1\}$ which is the same as the boolean value that b assigns to the variable x_j , and otherwise we set it to $EG_X(1)$.

Let $\tau_X^1 := \text{type}(EG_X(0)) = \text{type}(EG_X(1)) = (\ell_X^1, s_X^1)$ and $\tau_Y^1 := \text{type}(EG_Y(0)) = \text{type}(EG_Y(1)) = (\ell_Y^1, s_Y^1)$. Let $L_0 = (\ell_X^1 + \ell_Y^1)$ and note that $L_0 \leq c \cdot (2w + 2) \cdot D$ where D is some constant that upper bounds the lengths of our coordinate values, and therefore $L_0 = O(cw)$.

Let $\rho_T := \delta(\mathbf{0}_X, \mathbf{0}_Y) = \delta(\mathbf{0}_X, \mathbf{1}_Y) = \delta(\mathbf{1}_X, \mathbf{0}_Y)$ and $\rho_F := \delta(\mathbf{1}_X, \mathbf{1}_Y)$, and by definition we have $\rho_F \geq \rho_T + 1$. By definition of alignment gadgets there is a constant $C_1 \in \mathbb{Z}$ such that $\delta(EG_X(0), EG_Y(0)) = \delta(EG_X(1), EG_Y(0)) = \delta(EG_X(0), EG_Y(1)) = C_1 + (2w + 2)\rho_T =: \rho_1$ while $\delta(EG_X(1), EG_Y(1)) = C_1 + 2w\rho_F + 2\rho_T$ is larger than ρ_1 . Combining these formulas with the definitions of our gadgets proves the following claim.

Claim 3. *For any two vectors $a \in A, b \in B$ and two nodes u, v in the branching program, the δ -distance between the two gadgets $\text{RG}_X^{1,u \rightarrow v}(a)$ and $\text{RG}_Y^{1,u \rightarrow v}(b)$ is equal to ρ_1 if there is an edge from u to v in the branching program induced by the assignment (a, b) , and the δ -distance is larger otherwise.*

Level 1 index gadgets. For a boolean value $b \in \{0, 1\}$ we let $CV_X(b), CV_Y(b)$ be $\mathbf{0}_X, \mathbf{0}_Y$ respectively if $b = 0$ and $\mathbf{1}_X, \mathbf{1}_Y$ otherwise. Recall that $w = \log_2 W$, and for each number $z \in [W]$ we let $\bar{z} = (z_1, \dots, z_w) \in \{0, 1\}^w$ be the binary representation of z , and define the level 1 index gadgets as follows.

$$\text{IG}_X^1(z) := \text{GA}_X^{2w+2, \tau_Y}(CV_X(z_1), \dots, CV_X(z_w), CV_X(\neg z_1), \dots, CV_X(\neg z_w), \mathbf{0}_X, \mathbf{1}_X)$$

$$\text{IG}_Y^1(z) := \text{GA}_Y^{2w+2, \tau_X}(CV_Y(\neg z_1), \dots, CV_Y(\neg z_w), CV_Y(z_1), \dots, CV_Y(z_w), \mathbf{1}_Y, \mathbf{0}_Y)$$

Observe that by definition of our alignment gadgets, the sequence $\text{IG}_X^1(z)$ will have the same type $\text{type}(\text{IG}_X^1(z)) = \tau_X^1$ for all $z \in [W]$, and similarly $\text{type}(\text{IG}_Y^1(z)) = \tau_Y^1$ for all $z \in [W]$, which are the same types we had in the level-1 reachability gadgets. This definition allows us to prove the following property.

Claim 4. *For any $z, z' \in [W]$, the distance $\delta(\text{IG}_X^1(z), \text{IG}_Y^1(z'))$ is equal to ρ_1 if $z = z'$ and it is larger otherwise.*

Proof. First, assume that $z = z'$ and consider the alignment $A = \{(1, 1), \dots, (2w + 2, 2w + 2)\}$. In this alignment, $CV_X(\eta)$ is always aligned with $CV_Y(\neg \eta)$ and therefore the total cost is $\delta(\text{IG}_X^1(z), \text{IG}_Y^1(z')) \leq C_1 + (2w + 2) \cdot \rho_T = \rho_1$, which is also the minimum possible distance when aligning these sequences. Next, assume that $z \neq z'$ and note that in the alignment A as defined above there must be an aligned pair (i, i) in which the two aligned sequences are $\mathbf{1}_X$ and $\mathbf{1}_Y$, which implies that $\delta(A) \geq \rho_F + (2w + 1)\rho_T > 8\rho_T$. Also notice that any alignment except for A will leave some indices unaligned and will therefore cost at least $\rho_F + (2w + 1)\rho_T$ as well. Therefore, in this case $\delta(\text{IG}_X^1(z), \text{IG}_Y^1(z')) \geq C_1 + (2w + 1) \cdot \rho_T + \rho_F > \rho_1$. \square

Finally, we show that due to the last two coordinates, the distance between an index gadget and a reachability gadget is large.

Claim 5. *For any $z \in [W]$, and any two vectors $a \in A, b \in B$ and two nodes u, v in the branching program, we have that $\delta(\text{IG}_X^1(z), \text{RG}_Y^{1,u \rightarrow v}(b))$ and $\delta(\text{RG}_X^{1,u \rightarrow v}(a), \text{IG}_Y^1(z))$ are both larger than ρ_1 .*

Proof. The cost of any alignment other than $A = \{(1, 1), \dots, (2w + 2, 2w + 2)\}$ is at least $\rho_F + (2w + 1)\rho_T$ since it leaves some indices unaligned. The cost of A is also at least $\rho_F + (2w + 1)\rho_T$ since one of the last two pairs $((2w + 1), (2w + 1))$ or $((2w + 2), (2w + 2))$ will align $\mathbf{1}_X$ and $\mathbf{1}_Y$, by our construction. Therefore, the δ -distance is at least $C_1 + (2w + 1) \cdot \rho_T + \rho_F > \rho_1$. \square

This proves the base case for the following lemma, which is our main construction.

Lemma 7 (Reachability gadgets). *For all integers $k \geq 1$, the following statement is true: For any vectors $a \in A, b \in B$, integer $z, z' \in [W]$, nodes $u, v \in V$, we can construct gadgets:*

$$\text{RG}_X^{k-1,u \rightarrow v}(a) \in \mathcal{I}_{\tau_X^k}$$

$$\text{RG}_Y^{k-1,u \rightarrow v}(b) \in \mathcal{I}_{\tau_Y^k}$$

$$\text{IG}_X^{k-1}(z) \in \mathcal{I}_{\tau_X^k}$$

$$\text{IG}_Y^{k-1}(z') \in \mathcal{I}_{\tau_Y^k}$$

such that for some value ρ_k :

- $\delta(\text{RG}_X^{k,u \rightarrow v}(a), \text{RG}_Y^{k,u \rightarrow v}(b))$ is equal to ρ_k if there is a path of length 2^{k-1} from u to v in the branching program induced by the assignment (a, b) , and is larger otherwise.
- $\delta(\text{IG}_X^k(z), \text{IG}_Y^k(z'))$ is equal to ρ_k if $z = z'$ and is larger otherwise.
- $\delta(\text{IG}_X^k(z), \text{RG}_Y^{k,u \rightarrow v}(b))$, and $\delta(\text{RG}_X^{k,u \rightarrow v}(a), \text{IG}_Y^k(z'))$ are larger than ρ_k .
- $\tau_X^k = (\ell_X^k, s_X^k)$ and $\tau_Y^k = (\ell_Y^k, s_Y^k)$ only depend on k .
- The construction can be computed in $O(\ell_X + \ell_Y)$ time.
- The length of these gadgets can be upper bounded by $\ell_X^k, \ell_Y^k \leq (12W^2c^3)^k \cdot L_0$.

Proof. To prove the lemma, it remains to show the inductive step. Fix any $k > 1$ and from now on, assume that the statement of the lemma is true for $k - 1$, and we will show that it is also true for k .

The $k > 1$ Case. Let $u = (i, i_z) \in [T] \times [W]$ be a node on layer i and let $v = (j, j_z) \in [T] \times [W]$ be a node on layer j . Assume that $j - i = 2^{k-1}$ and we are at level $k \in [\log_2 T]$ of the recursive construction. We define $h = \frac{i+j}{2}$ to be the layer in the middle between i and j and note that $h - i = j - h = 2^{k-2}$. For each node $w = (h, z)$ for $z \in [W]$ in layer h we will add a gadget that enables the path from u to v to pass through this node w . We do this by recursively adding the two gadgets $\text{RG}^{k-1, u \rightarrow w}$ and $\text{RG}^{k-1, w \rightarrow v}$.

To do this formally, we will combine an alignment gadget with an OR gadget.

We start by defining *path gadgets*, which will be used to determine whether there is a path from u to v through a specific node $w = (h, z)$. For all vectors $a \in A, b \in B$ and all numbers $z \in [W]$, we define:

$$\text{PG}_X^{k, u \rightarrow v}(a, z) := \text{GA}_X^{3, \tau_Y^{k-1}} \left(\text{RG}_X^{k-1, u \rightarrow (h, z)}(a), \text{RG}_X^{k-1, (h, z) \rightarrow v}(a), \text{IG}_X^{k-1}(z) \right)$$

$$\text{PG}_Y^{k, u \rightarrow v}(b, z) := \text{GA}_Y^{3, \tau_X^{k-1}} \left(\text{RG}_Y^{k-1, u \rightarrow (h, z)}(b), \text{RG}_Y^{k-1, (h, z) \rightarrow v}(b), \text{IG}_Y^{k-1}(z) \right)$$

and note that the types $\tau_X^{(1), k} := \text{type}(\text{PG}_X^{k, u \rightarrow v}(a, z)) = (\ell_X^{(1), k}, s_X^{(1), k})$ and $\tau_Y^{(1), k} := \text{type}(\text{PG}_Y^{k, u \rightarrow v}(b, z)) = (\ell_Y^{(1), k}, s_Y^{(1), k})$ depend only k , and that $\ell_X^{(1), k}, \ell_Y^{(1), k} \leq c \cdot 3 \cdot (\ell_X^{k-1} + \ell_Y^{k-1}) \leq 3c \cdot 2 \cdot (12W^2c^3)^{k-1} \cdot L_0$.

Claim 6. *There is a constant $C_k^{(1)} \in \mathbb{Z}$ such that for all vectors $a \in A, b \in B$ and integers $z, z' \in [W]$ we have that*

$$\delta(\text{PG}_X^{k, u \rightarrow v}(a, z), \text{PG}_Y^{k, u \rightarrow v}(b, z')) = C_k^{(1)} + 3 \cdot \rho_{k-1}$$

if $z = z'$ and there is a path from u to v through (h, z) in the branching program induced by (a, b) , and the δ -distance is larger otherwise.

Proof. First, observe that the minimal cost of any alignment $A \in \mathcal{A}_{3,3}$ is $3 \cdot \rho_{k-1}$, regardless of a, b, z, z' . Next, observe that any alignment other than $A = \{(1, 1), (2, 2), (3, 3)\}$ will have cost at least $3 \cdot \rho_{k-1} + 1$. This follows from the definition of the cost of an alignment, because when aligning an index gadget to a reachability gadget we incur a cost of at least $\rho_{k-1} + 1$, and therefore leaving any index unaligned will incur at least this cost. Finally, we focus on the alignment A and compute its cost.

We argue that $\text{cost}(A) = 3\rho_{k-1}$ if a, b, z, z' satisfy the statement in the claim, and $\text{cost}(A)$ is larger otherwise. If $z \neq z'$ then the pair $(3, 3)$ incurs a cost of $\rho_{k-1} + 1$ and we are done. Otherwise, $z = z'$, and we have that

$$\text{cost}(A) = \rho_{k-1} + \delta(\text{RG}_X^{k-1, u \rightarrow (h, z)}(a), \text{RG}_Y^{k-1, u \rightarrow (h, z)}(b)) + \delta(\text{RG}_X^{k-1, (h, z) \rightarrow v}(a), \text{RG}_Y^{k-1, (h, z) \rightarrow v}(b)).$$

By our inductive hypothesis, the last two summands are equal to ρ_{k-1} if u can reach (h, z) and (h, z) can reach v in the branching program induced by (a, b) , while at least one of them is larger otherwise. Therefore, $\text{cost}(A) = 3 \cdot \rho_{k-1}$ if there is a path from u to v through (h, z) in the branching program induced by (a, b) and is larger otherwise. By definition of alignment gadget, there is a constant $C_k^{(1)}$ such that the statement of the claim holds. \square

We are now ready to define our reachability gadgets, using OR gadgets. Recall that by Lemma 3, any measure that admits alignment gadgets of size cn also admits OR gadgets of size c^2n^2 .

$$\text{RG}_X^{k, u \rightarrow v}(a) := \text{OR}_X^{W, \tau_Y^{(1), k}} \left(\text{PG}_X^{k, u \rightarrow v}(a, 1), \dots, \text{PG}_X^{k, u \rightarrow v}(a, W) \right)$$

$$\text{RG}_Y^{k,u \rightarrow v}(b) := \text{OR}_Y^{W, \tau_X^{(1),k}} \left(\text{PG}_Y^{k,u \rightarrow v}(b, 1), \dots, \text{PG}_Y^{k,u \rightarrow v}(b, W) \right)$$

Note that the types $\tau_X^k := \text{type}(\text{RG}_X^{k,u \rightarrow v}(a)) = (\ell_X^k, s_X^k)$ and $\tau_Y^k := \text{type}(\text{RG}_Y^{k,u \rightarrow v}(b)) = (\ell_Y^k, s_Y^k)$ depend only k , and that

$$\ell_X^k, \ell_Y^k \leq c^2 \cdot W^2 \cdot (\ell_X^{(1),k} + \ell_Y^{(1),k}) \leq c^2 W^2 \cdot (2 \cdot 6c \cdot (12W^2 c^3)^{k-1} \cdot L_0) = (12W^2 c^3)^k \cdot L_0.$$

Claim 7. *There is a constant $\rho_k \in \mathbb{Z}$ such that for all vectors $a \in A, b \in B$ we have that*

$$\delta(\text{RG}_X^{k,u \rightarrow v}(a), \text{RG}_Y^{k,u \rightarrow v}(b)) = \rho_k$$

there is a path from u to v in the branching program induced by (a, b) , and the δ -distance is larger otherwise.

Proof. By definition of OR gadgets, we know that there is a constant $C_k^{(2)} \in \mathbb{Z}$ such that

$$\delta(\text{RG}_X^{k,u \rightarrow v}(a), \text{RG}_Y^{k,u \rightarrow v}(b)) = C_k^{(2)} + \min_{z, z' \in [W]} \delta(\text{PG}_X^{k,u \rightarrow v}(a, z), \text{PG}_X^{k,u \rightarrow v}(b, z')).$$

By Claim 6, there is a constant $C_k^{(1)}$ such that $\min_{z, z' \in [W]} \delta(\text{PG}_X^{k,u \rightarrow v}(a, z), \text{PG}_X^{k,u \rightarrow v}(b, z')) \geq C_k^{(1)} + 3\rho_{k-1}$ with equality if and only if for some $z \in [W]$ there is a path from u to v through (h, z) in the branching program induced by (a, b) . Therefore, for $\rho_k := C_k^{(2)} + C_k^{(1)} + 3\rho_{k-1}$ the statement of the claim holds. \square

To complete the proof of Lemma 7 we need to construct index gadgets. These gadgets have straightforward functionality, but their definitions are a bit complicated because we must enforce that the type of these gadgets is exactly the same as the types of the reachability gadgets.

For all $z \in [W]$ we define

$$\text{IG}_X^{(1),k}(z) := \text{GA}_X^{3, \tau_Y^{k-1}}(\text{IG}_X^{k-1}(z), \text{IG}_X^{k-1}(z), \text{IG}_X^{k-1}(z))$$

$$\text{IG}_Y^{(1),k}(z) := \text{GA}_Y^{3, \tau_X^{k-1}}(\text{IG}_Y^{k-1}(z), \text{IG}_Y^{k-1}(z), \text{IG}_Y^{k-1}(z))$$

so that the types are $\tau_X^{(1),k}$ and $\tau_Y^{(1),k}$. By an argument similar (but simpler) to the one in Claim 6, we get that for all integers $z, z' \in [W]$ we have $\delta(\text{IG}_X^{(1),k}(z), \text{IG}_X^{(1),k}(z')) = C_k^{(1)} + 3 \cdot \rho_{k-1}$ if $z = z'$ and the δ -distance is larger otherwise. Another straightforward consequence of these definitions is that for any vectors $a \in A, b \in B$ and integers $z, z' \in [W]$ we have that $\delta(\text{IG}_X^{(1),k}(z), \text{PG}_Y^{k,u \rightarrow v}(b, z'))$ and $\delta(\text{PG}_X^{k,u \rightarrow v}(a, z'), \text{IG}_Y^{(1),k}(z))$ are larger than $C_k^{(1)} + 3 \cdot \rho_{k-1}$.

Finally, we define

$$\text{IG}_X^k(z) := \text{OR}_X^{W, \tau_Y^{(1),k}} \left(\text{IG}_X^{(1),k}(z), \dots, \text{IG}_X^{(1),k}(z) \right)$$

$$\text{IG}_Y^k(z) := \text{OR}_Y^{W, \tau_X^{(1),k}} \left(\text{IG}_Y^{(1),k}(z), \dots, \text{IG}_Y^{(1),k}(z) \right)$$

so that the types are τ_X^k and τ_Y^k . Again, by an argument similar (but simpler) to the one in Claim 7, we have that $\delta(\text{IG}_X^k(z), \text{IG}_Y^k(z')) = \rho_k$ if $z = z'$ and is larger otherwise. And, moreover, that $\delta(\text{IG}_X^k(z), \text{RG}_Y^k(b)), \delta(\text{RG}_X^k(a), \text{IG}_Y^k(z))$ are larger than ρ_k , for all a, b, z .

To complete the proof of Lemma 7, we remark that this construction takes linear time in its output, since the runtime is dominated by the constructions of alignment and OR gadgets. \square

We now continue with the reduction from BP-SAT to the problem of computing the δ -similarity of two sequences.

Recall that $t = \log_2 T$, and let the start node of the branching program be $u_{\text{start}} = (1, 1)$ and the only accept node be $u_{\text{acc}} = (2^t, 1)$. Intuitively, we would like to define vector gadgets of the form $VG(a) = \text{RG}^{t, u_{\text{start}} \rightarrow u_{\text{acc}}}(a)$ and $VG(b) = \text{RG}^{t, u_{\text{start}} \rightarrow u_{\text{acc}}}(b)$ so that $\delta(VG(a), VG(b))$ will tell us whether (a, b) is a satisfying pair or not (whether it induces a path from the start node to the accepting node). However, this does not quite work for the following technical reason: When we combine all these $2n$ vector gadgets into two sequences x, y , the score of $\delta(VG(a'), VG(b'))$ of other pairs a', b' will affect the overall score, and could potentially hide the contribution of the satisfying pair.

To fix this, we “normalize” the vector gadgets so that the distance $\delta(VG(a'), VG(b'))$ of unsatisfying pairs is fixed (and is slightly worse than the distance of satisfying pairs). This “normalization” trick was introduced by Backurs and Indyk in their reduction from OV to Edit-Distance.

We will need the following simple constructions of instances S^k such that the distance between S^k and any reachability gadget is fixed, and is slightly worse than the score of a “good pair”.

Claim 8. *For all $k \geq 1$, there are sequences $S^k, T^k \in \mathcal{I}_{\tau_X^k}$ such that for all vectors $b \in B$ and nodes u, v we have that $\delta(S^k, \text{RG}_Y^{k, u \rightarrow v}) = \rho_k + (\rho_F - \rho_T)$ and $\delta(T^k, \text{RG}_Y^{k, u \rightarrow v}) = \rho_k$.*

Proof. Base case, $k = 1$: we define

$$S^1 := \text{GA}_X^{2w+2, \tau_Y}(\mathbf{0}_X, \dots, \mathbf{0}_X, \mathbf{1}_X, \mathbf{1}_X)$$

$$T^1 := \text{GA}_X^{2w+2, \tau_Y}(\mathbf{0}_X, \dots, \mathbf{0}_X, \mathbf{0}_X, \mathbf{0}_X)$$

which ensures that $\text{type}(S^1) = \text{type}(T^1) = \tau_X^1$ and that $\delta(S^1, \text{RG}_Y^{1, u \rightarrow v}(b)) = C_1 + 7\rho_T + \rho_F = \rho_1 + (\rho_F - \rho_T)$, while $\delta(T^1, \text{RG}_Y^{1, u \rightarrow v}(b)) = \rho_1$.

Inductive step: assume the statement holds for $k - 1$ for some sequence $S^{k-1} \in \tau_X^{k-1}$ and we will prove it for k . For all $z \in [W]$, consider the sequences:

$$S^{(1),k}(z) := \text{GA}_X^{3, \tau_Y^{k-1}}(T^{k-1}, S^{k-1}, \text{IG}_X^{k-1}(z)) \in \mathcal{I}_{\tau_X^{(1),k}}$$

$$T^{(1),k}(z) := \text{GA}_X^{3, \tau_Y^{k-1}}(T^{k-1}, T^{k-1}, \text{IG}_X^{k-1}(z)) \in \mathcal{I}_{\tau_X^{(1),k}}$$

and then define:

$$S^k := \text{OR}_X^{W, \tau_Y^{(1),k}}(S^{(1),k}(1), \dots, S^{(1),k}(W)) \in \mathcal{I}_{\tau_X^k}$$

$$T^k := \text{OR}_X^{W, \tau_Y^{(1),k}}(T^{(1),k}(1), \dots, T^{(1),k}(W)) \in \mathcal{I}_{\tau_X^k}$$

And the claim follows by simple calculations. □

We can now define our *normalized vector gadgets*. For all vectors $a \in A, b \in B$ we define:

$$\text{NVG}_X(a) := \text{GA}_X^{1, \tau_Y^t}(S^t, \text{RG}_X^{t, u_{\text{start}} \rightarrow u_{\text{acc}}}(a))$$

$$\text{NVG}_Y(b) := \text{GA}_Y^{2, \tau'_X}(\text{RG}_Y^{t, u_{\text{start}} \rightarrow u_{\text{acc}}}(b))$$

We denote the types of these gadgets by $\text{type}(\text{NVG}_X(a)) =: \tau'_X$ and $\text{type}(\text{NVG}_Y(b)) =: \tau'_Y$ and remark that they are independent of a, b . Also, note that the length of these gadgets can be upper bounded by

$$c \cdot 2 \cdot (\ell_X^t + \ell_Y^t) \leq 2c \cdot (12W^2c^3)^t \cdot L_0 = O((12W^2c^3)^{\log_2 T} \cdot c \log_2 W).$$

Lemma 8 (Vector Gadgets). *There is a constant $C \in \mathbb{Z}$ such that for any two vectors $a \in A, b \in B$ we have that:*

$$\delta(\text{NVG}_X(a), \text{NVG}_Y(b)) = C + \rho_t$$

if the pair (a, b) satisfies the branching program, and otherwise

$$\delta(\text{NVG}_X(a), \text{NVG}_Y(b)) = C + \rho_t + (\rho_F - \rho_T).$$

Proof. The proof follows from the definition of alignment gadgets and from Lemma 7 and Claim 8. \square

Let $A = \{a_1, \dots, a_{2^{n/2}}\}$ and $B = \{b_1, \dots, b_{2^{n/2}}\}$ be our sets of vectors. Our final sequences are defined as follows:

$$x := \text{GA}_X^{2^{n/2}, \tau'_Y}(\text{NVG}_X(a_1), \dots, \text{NVG}_X(a_{2^{n/2}}), \text{NVG}_X(a_1), \dots, \text{NVG}_X(a_{2^{n/2}}))$$

$$y := \text{GA}_Y^{2 \cdot 2^{n/2}, \tau'_X}(\text{NVG}_Y(b_1), \dots, \text{NVG}_Y(b_{2^{n/2}}))$$

First, we upper bound the length of these sequences:

$$|x|, |y| \leq c \cdot 2 \cdot 2^{n/2} \cdot O((12W^2c^3)^{\log_2 T} \cdot c \log_2 W) = O((12W^2c^3)^{\log_2 T} \cdot 2^{n/2} \cdot c^2 \log_2 W) = T^{O(\log W)} \cdot 2^{n/2}$$

Finally, the theorem follows from this claim which shows that the answer to our BP-SAT can be deduced from $\delta(x, y)$.

Claim 9. *There is a constant $C^* \in \mathbb{Z}$ such that*

$$\delta(x, y) \leq C^* + (2^{n/2} - 1) \cdot (C + \rho_t + (\rho_F - \rho_T)) + (C + \rho_t)$$

if and only if there is a pair $a \in A, b \in B$ that satisfies the branching program.

Proof. If there is no satisfying pair, then the cost of any alignment $A \in \mathcal{A}_{2^{n/2}, 2 \cdot 2^{n/2}}$ is equal to $2^{n/2}$ summands all of which are at least $C + \rho_t + (\rho_F - \rho_T)$, and therefore $\delta(x, y) \geq C^* + 2^{n/2} \cdot (C + \rho_t + (\rho_F - \rho_T))$.

On the other hand, if $a_i \in A, b_j \in B$ is a satisfying pair, for some $i, j \in [2^{n/2}]$, consider the structured alignment $A = \{(\Delta + 1, 1), \dots, (\Delta + 2^{n/2}, 2^{n/2})\}$ where $\Delta := i - j$ if $i \geq j$ and $\Delta := 2^{n/2} + i - j$ otherwise. The cost of A is equal to $2^{n/2}$ summands, one of which is $\delta(\text{NVG}_X(a_i), \text{NVG}_Y(b_j)) \leq (C + \rho_t)$, while the others are at most $(C + \rho_t + (\rho_F - \rho_T))$, which implies that

$$\delta(x, y) \leq \text{cost}(A) \leq C^* + (2^{n/2} - 1) \cdot (C + \rho_t + (\rho_F - \rho_T)) + (C + \rho_t).$$

\square

\square

6 Faster SAT implies Circuit Lower Bounds

A direct corollary of our reduction (Theorem 2) is that faster Edit Distance or LCS implies faster Branching-Program-SAT.

Corollary 3. *If Edit Distance or LCS on two binary sequences of length N can be solved in $O(N^2/t(N))$ time, then SAT on BPs on n variables of width W and length T can be solved in $(2^n \cdot T^{O(\log W)})/t(2^{n/2} \cdot T^{O(\log W)})$ time.*

Here, we show how we get our corollaries about surprisingly strong circuit lower bounds from such faster BP-SAT algorithm.

We will apply the following generic results, which follow from the literature:

Theorem 7 ([44]). *Suppose there is a satisfiability algorithm for bounded fan-in formulas of size n^k running in $O(2^n/n^k)$ time, for all constants $k > 0$. Then $\text{NTIME}[2^{O(n)}]$ is not contained in non-uniform NC^1 .*

Theorem 7 follows directly from the reference [44].

Theorem 8 ([44, 46]). *Let $n \leq S(n) \leq 2^{o(n)}$ be time constructible and monotone non-decreasing. Let \mathcal{C} be a class of circuits. Suppose there is an SAT algorithm for n -input circuits which are arbitrary functions of three $O(S(n))$ -size circuits from \mathcal{C} , that runs in $O(2^n/(n^{10} \cdot S(n)))$ time. Then E^{NP} does not have $S(n)$ -size circuits.*

Alternatively, suppose there is an SAT algorithm for n -input circuits which are ANDs of $O(S(n))$ arbitrary functions of three $O(S(n))$ -size circuits from \mathcal{C} , that runs in $O(2^n/n^{10})$ time. Then the same circuit lower bound holds.

Proof. (Sketch) The argument essentially follows [44], with some minor modifications. First, assuming E^{NP} has $S(n)$ -size circuits, it follows that every verifier V for every $L \in \text{NTIME}[2^n]$ has $S(n)$ -size “witness circuits” which can print the bits of a witness to an arbitrary instance x of L .

Towards a contradiction, we want to simulate every $L \in \text{NTIME}[2^n]$ in nondeterministic time $o(2^n)$; this will contradict the nondeterministic time hierarchy [48]. On an input x of length n , we begin by reducing x to an instance C_x of the Succinct3SAT problem, in $\text{poly}(n)$ time [44]. We have the property that $x \in L$ if and only if the truth table of C_x encodes a satisfiable 3CNF formula of at most $2^n \cdot n^4$ size.

Now, in E^{NP} , on an input (x, i, j) (where $i \leq n + 4 \log n$ and $j \leq n^4$ are binary integers), we can (a) construct the circuit C_x , (b) compute a canonical (lexicographically first) witness circuit D of $S(n)$ -size for the instance C_x , (c) construct (in $\text{poly}(S(n))$ time) an $O(S(n))$ -size circuit D' which is UNSAT if and only if D encodes a valid witness to C_x , and (d) evaluate $D'(i)$, and output the value of the j th gate of D . By assumption, the language L' of such triples (x, i, j) has $S(n)$ -size circuits.

The next step of our simulation is to guess the circuit D of the previous paragraph (constructing the circuit D' using D and C_x) as well as a circuit E of size $O(S(n))$ implementing the language L' of triples, for the case of $|x| = n$. To verify whether C_x encodes a satisfiable formula, we construct circuits F_j for all gates $j = 1, \dots, O(n^4)$ of D' . The circuit F_j takes an input $i \leq n + 4 \log n$, and uses a constant number of AND/OR gates to check that the inputs j_1, j_2 to the j th gate of $D'(i)$ match the output, using the information provided by $E(x, i, j_1)$, $E(x, i, j_2)$, and $E(x, i, j)$ (which are supposed to print exactly these outputs). That is, each F_j is a function of three copies of $E(x, i, j)$, each of which have size $O(S(n))$.

Therefore each F_j has size $O(S(n))$, and by assumption we can compute for each F_j whether it is true on all of its $2^n \cdot n^4$ inputs, in time $O(2^n/(n^5 \cdot S(n)))$. Over all $O(S(n))$ different F_j , these checks cost only $O(2^n/n^5)$ time in total. If all checks pass (meaning that $E(x, i, j)$ reports the correct gate value for all (x, i, j)), we finally examine whether D' is true on all of its inputs – this can be done by calling UNSAT on $\neg D'$ or on $\neg E(x, i, j^*)$ where j^* is the index of the output gate of D' . The circuit D' is true on all of its inputs if and only if the circuit D is a witness for C_x ; hence we *accept* if and only if all of the above checks pass and D' is true on all inputs.

Alternatively, suppose there is an SAT algorithm for n -input circuits which are ANDs of $O(S(n))$ arbitrary functions of three $O(S(n))$ -size circuits from \mathcal{C} , that runs in $O(2^n/n^{10})$ time. Then we can check all of the aforementioned F_j circuits simultaneously, by taking an AND of them. This AND has fan-in $O(S(n))$, and the resulting circuit has exactly the form for which we are assumed to have an efficient SAT algorithm. The rest of the argument is the same. \square

Boolean formulas (without further restrictions) are trivially closed under the AND and OR as prescribed in Theorem 8. It is also easy to see that nondeterministic branching programs are closed under AND and OR: to simulate an OR of several nondeterministic BPs, make a new start node that connects to all the start nodes of the BPs, without reading a variable. To simulate an AND of BPs, connect the collection of branching programs in sequence, with the accept node of the previous branching program connected to the start node of the next branching program.

Consider Corollary 1 from the introduction:

Reminder of Corollary 1 *If Edit Distance or LCS on two binary sequences of length N is in $O(N^{2-\varepsilon})$ time for some $\varepsilon > 0$, then the complexity class $\mathbf{E}^{\mathbf{NP}}$ does not have:*

1. *non-uniform $2^{o(n)}$ -size Boolean formulas,*
2. *non-uniform $o(n)$ -depth circuits of bounded fan-in, and*
3. *non-uniform $2^{o(n^{1/2})}$ -size nondeterministic branching programs.*

Furthermore, $\mathbf{NTIME}[2^{O(n)}]$ is not in non-uniform NC.

We will explain how to obtain the corollary from Theorems 8,7, and our Theorem 2. Items 1 and 2 follow from combining the SAT-to-lower-bounds connection (Theorem 8) with the $O(2^{n-\varepsilon n/2})$ -time algorithm for SAT on $2^{o(n)}$ -size formulas, implied by Theorem 2. Item 3 follows from applying the same connection to the $O(2^{n-\varepsilon n/2})$ -time algorithm for SAT on nondeterministic branching programs of $2^{o(\sqrt{n})}$ size, implied by Theorem 2. Finally, the conclusion “ $\mathbf{NTIME}[2^{O(n)}]$ is not in non-uniform NC” follows from combining the $O(2^{n-\varepsilon n/2})$ -time algorithm for SAT on $o(n)$ -depth circuits and Theorem 7.

Even sufficiently large $\text{poly}(\log n)$ improvements in algorithms for Edit-Distance would establish strong new lower bounds, as in Corollary 2:

Reminder of Corollary 2 *If Edit Distance or LCS on two binary sequences of length N can be solved in $O(n^2/\log^c n)$ time for every $c > 0$, then $\mathbf{NTIME}[2^{O(n)}]$ does not have non-uniform polynomial-size log-depth circuits.*

The corollary follows from the fact that the hypothesis implies a BP-SAT algorithm for n^c -size $O(1)$ -width branching programs running in $O(2^n/n^c)$ time for all constants $c > 0$ (Theorem 2). By standard reductions from log-depth circuits to formulas, and Barrington’s reduction from Boolean formulas to branching programs, we obtain an analogous algorithm for Formula-SAT, which by the SAT-to-lower-bounds connection (Theorem 7) implies the circuit lower bound.

Acknowledgement. We thank Arturs Backurs for comments on an earlier version.

References

- [1] <https://www.bostonglobe.com/ideas/2015/08/10/computer-scientists-have-looked-for-solution-tX00qNRnbKrClfUPmavifK/story.html>.
- [2] <https://www.quantamagazine.org/20150929-edit-distance-computational-complexity/>.
- [3] <http://news.mit.edu/2015/algorithm-genome-best-possible-0610>.
- [4] <http://www.scottaaronson.com/blog/?p=2335>.
- [5] <https://rjlipton.wordpress.com/2015/06/01/puzzling-evidence/>.
- [6] A. Abboud, A. Backurs, T. Hansen, V. Vassilevska Williams, and O. Zamir. Subtree isomorphism revisited. In *Proc. of the 27th SODA, to appear*, 2016.
- [7] A. Abboud, A. Backurs, and V. Vassilevska Williams. Tight Hardness Results for LCS and other Sequence Similarity Measures. In *FOCS*, 2015.
- [8] A. Abboud, F. Grandoni, and V. V. Williams. Subcubic equivalences between graph centrality problems, APSP and diameter. In *Proc. of the 26th SODA*, pages 1681–1697, 2015.
- [9] A. Abboud and V. Vassilevska Williams. Popular conjectures imply strong lower bounds for dynamic problems. *FOCS*, 2014.
- [10] A. Abboud, V. Vassilevska Williams, and O. Weimann. Consequences of faster sequence alignment. *ICALP*, 2014.
- [11] A. Abboud, V. Vassilevska Williams, and H. Yu. Matching triangles and basing hardness on an extremely popular conjecture. In *Proc. of the 47th STOC*, pages 41–50, 2015.
- [12] A. Abboud, R. Williams, and H. Yu. More applications of the polynomial method to algorithm design. In *SODA*, 2015.
- [13] A. Abboud, V. V. Williams, and J. R. Wang. Approximation and fixed parameter subquadratic algorithms for radius and diameter. In *Proc. of the 27th SODA, to appear*, 2016.
- [14] J. Alman and R. Williams. Probabilistic polynomials and hamming nearest neighbors. *arXiv preprint arXiv:1507.05106*, 2015.
- [15] S. Arora and B. Barak. *Computational complexity: a modern approach*. Cambridge University Press, 2009.
- [16] A. Backurs and P. Indyk. Edit Distance Cannot Be Computed in Strongly Subquadratic Time (unless SETH is false). In *STOC*, 2015.
- [17] N. Bansal and R. Williams. Regularity lemmas and combinatorial algorithms. In *Proc. of the 50th FOCS*, pages 745–754, 2009.

- [18] D. A. Barrington. Bounded-width polynomial-size branching programs recognize exactly those languages in NC1. *Journal of Computer and System Sciences*, 38(1):150–164, 1989.
- [19] C. Beck and R. Impagliazzo. Strong ETH holds for regular resolution. In *Proc. of the 45th STOC*, pages 487–494, 2013.
- [20] H. Bodlaender, R. G. Downey, M. Fellows, and H. T. Wareham. The parameterized complexity of sequence alignment and consensus. In *Combinatorial Pattern Matching*, pages 15–30, 1994.
- [21] K. Bringmann. Why walking the dog takes time: Frechet distance has no strongly subquadratic algorithms unless seth fails. *FOCS*, 2014.
- [22] K. Bringmann and M. Kunnemann. Quadratic Conditional Lower Bounds for String Problems and Dynamic Time Warping. In *FOCS*, 2015.
- [23] K. Bringmann and W. Mulzer. Approximability of the Discrete Fréchet Distance. In *31st International Symposium on Computational Geometry (SoCG 2015)*, pages 739–753, 2015.
- [24] C. Calabro, R. Impagliazzo, and R. Paturi. A duality between clause width and clause density for SAT. In *Proc. of the 21st CCC*, pages 252–260, 2006.
- [25] C. Calabro, R. Impagliazzo, and R. Paturi. The complexity of satisfiability of small depth circuits. In *Proc. of the 4th IWPEC*, pages 75–85, 2009.
- [26] T. M. Chan. Speeding up the four russians algorithm by about one more logarithmic factor. In *Proc. of the 26th SODA*, pages 212–217, 2015.
- [27] T. M. Chan and R. Williams. Deterministic apsp, orthogonal vectors, and more: Quickly derandomizing the polynomial method. In *Proc. of the 27th SODA*, 2016.
- [28] V. Chvatal, D. Klarner, and D. E. Knuth. Selected combinatorial research problems. Technical Report STAN-CS-72-292, Computer Science Department, Stanford University.
- [29] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms, Third Edition*. The MIT Press, 3rd edition, 2009.
- [30] E. Dantsin and E. A. Hirsch. Worst-case upper bounds. In *Handbook of Satisfiability*, pages 403–424. 2009.
- [31] R. Impagliazzo and R. Paturi. On the complexity of k-sat. *Journal of Computer and System Sciences*, 62(2):367–375, 2001.
- [32] R. Impagliazzo, R. Paturi, and F. Zane. Which problems have strongly exponential complexity? *Journal of Computer and System Sciences*, 63:512–530, 2001.
- [33] H. Jahanjou, E. Miles, and E. Viola. Local reductions. In *Proc. of the 42nd ICALP*, pages 749–760, 2015.
- [34] A. G. Jørgensen and S. Pettie. Threesomes, degenerates, and love triangles. In *Proc. of the 55th FOCS*, pages 621–630, 2014.

- [35] W. J. Masek and M. S. Paterson. A faster algorithm computing string edit distances. *Journal of Computer and System sciences*, 20(1):18–31, 1980.
- [36] M. Patrascu and R. Williams. On the possibility of faster sat algorithms. *SODA*, 2010.
- [37] R. Paturi, P. Pudlák, M. E. Saks, and F. Zane. An improved exponential-time algorithm for k -sat. *J. ACM*, 52(3):337–364, 2005.
- [38] K. Pietrzak. On the parameterized complexity of the fixed alphabet shortest common supersequence and longest common subsequence problems. *Journal of Computer and System Sciences*, 67(4):757–771, 2003.
- [39] L. Roditty and V. Vassilevska Williams. Fast approximation algorithms for the diameter and radius of sparse graphs. *STOC*, 2013.
- [40] W. J. Savitch. Relationships between nondeterministic and deterministic tape complexities. *Journal of computer and system sciences*, 4(2):177–192, 1970.
- [41] P. M. Spira. On time-hardware complexity tradeoffs for boolean functions. In *Proceedings of the 4th Hawaii Symposium on System Sciences*, pages 525–527, 1971.
- [42] L. G. Valiant. *Graph-theoretic arguments in low-level complexity*. Springer, 1977.
- [43] R. Williams. A new algorithm for optimal 2-constraint satisfaction and its implications. *Theoretical Computer Science*, 348(2):357–365, 2005.
- [44] R. Williams. Improving exhaustive search implies superpolynomial lower bounds. *SIAM J. Comput.*, 42(3):1218–1244, 2013.
- [45] R. Williams. Faster all-pairs shortest paths via circuit complexity. In *Proc. of 46th STOC*, pages 664–673, 2014.
- [46] R. Williams. Nonuniform ACC circuit lower bounds. *J. ACM*, 61(1):2:1–2:32, 2014.
- [47] H. Yu. An improved combinatorial algorithm for boolean matrix multiplication. In *Proc. of the 42nd ICALP*, pages 1094–1105, 2015.
- [48] S. Žák. A turing machine time hierarchy. *Theoretical Computer Science*, 26(3):327–333, 1983.